



University
of Glasgow

<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

The development of experimental and computational tools for studying
protein interactions:
Applied to the bacterial toxin pneumolysin.

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Glasgow

by

Nithin Sudharam Rai BSc (London). MSc (Leicester)
Division of Infection and Immunity,
Institute of Biomedical and Life Sciences
University of Glasgow

July 2001

ProQuest Number: 10646971

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10646971

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT

Much attention has been given to the study of protein interactions, whether with other proteins (of the same or different type), with nucleic acids or with cell membranes. Such interactions can be understood by using structural models and by biophysical monitoring of the kinetics and structural basis of the interaction. Pneumolysin is a 53 kDa bacterial protein toxin that inserts into eukaryotic membranes where it self-associates to form pores in the membrane. This process involves the initial interaction between toxin and membrane, followed by interaction between toxin monomers. Because pneumolysin elicits an immune response, it also interacts with antibodies. It therefore represents a system that is involved in a variety of different interactions. The aim of this thesis was to develop modelling and experimental methodologies that would be used to further understand the complex behaviour of pneumolysin and its behaviour with other structures. The choice of methodologies were based on those used by Gilbert (1998) i.e. bead modelling and haemolytic studies. This thesis describes: the development and testing of the bead model generating program, *SOMO*; a purification protocol that used perfusion chromatography to purify pneumolysin, and an electrochemical instrument (Enzymometer) for monitoring haemolytic activity.

SOMO was successfully tested on four experimentally characterised proteins: bovine pancreatic trypsin inhibitor, ribonuclease A, lysozyme and the much larger, dimeric protein, citrate synthase. The results for the theoretical hydrodynamic predictions for these models show a very good correlation with experimental values particularly when water volumes of 24 Å were used for modelling the hydration layer. This program is now ready to be tested on experimental systems such as pneumolysin. In preparation for this, pneumolysin was successfully purified (in a matter of minutes) by perfusion chromatography. While developing a continuous assaying method for pneumolysin using the enzymometer, it was shown that maximum toxin activity for both rabbit and sheep erythrocytes (at 37 °C, in the presence of either magnesium or calcium divalent ions) occurred at approximately 0.4 mM Ca^{++} and at concentrations beyond 0.2 mM Mg^{++} . Below these concentrations caused a decrease in toxin activity. Further studies also indicated a difference between ion leakage and haemoglobin leakage which suggests that cell lysis might occur by osmotic shock; a proposition thought unlikely based on previous studies (Blumenthal and Habig, 1984; Duncan, 1974).

DECLARATION

This thesis is the original work of the author except where otherwise stated. The accompanying compact disc contains material written by the author that can be found in the directory named 'somo'. Other programs used in this project have also been included in the CD, and their use will require the permission from the appropriate authors.

Nithin Rai

July 2001

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION 2

1.1 Characterising protein structure and interactions 2

1.2 The present study 7

CHAPTER 2

EXPERIMENTAL DETERMINATION OF SOLUTION PARAMETERS 9

2.1 Small angle scattering 9

2.1.1 Basic Principles 10

2.1.2 Deriving physical parameters from small angle scattering 13

2.2 Analytical ultracentrifugation 15

2.2.1 Sedimentation velocity 15

2.2.2 Sedimentation equilibrium 17

CHAPTER 3

THEORETICAL DETERMINATION OF SOLUTION PARAMETERS 20

3.1 Whole body approach 20

3.2 Multibody approach 23

3.2.1 Bead-filled models 24

3.2.2 Shell models 26

3.3 Hydrodynamic treatment for rigid bodies 28

3.4 Scattering properties from bead models 31

CHAPTER 4

ANALYSIS AND IMPLEMENTATION OF SOFTWARE REQUIREMENTS 36

4.1	Program design	37
4.2	Algorithm optimisation	40
4.2.1	Bead model generation	42
4.2.2	Overlap removal	47
4.2.3	Outward translation	51
4.2.4	Bead popping	56
4.2.5	Look-up data table	61
4.3	Graphical interfaces	64
4.3.1	Availability of source code	66
4.3.2	Types of operating systems	66
4.3.3	Basic manipulations of models	66
4.3.4	Graphics and visualisation	67
4.3.5	Display features	67
4.3.6	Additional features	68

CHAPTER 5

TESTING THE ALGORITHMS ON WELL CHARACTERISED PROTEINS USING HYDRODYNAMIC THEORY 69

5.1	Introduction	69
5.2	Method	70
5.2.1	ASA	71
5.2.2	SOMO	73
5.2.3	RAZ	75
5.2.4	SUPC	75

5.3	Results	76
5.3.1	Bead models tested at different water volumes and at different resolutions	76
5.3.2	Effect of buried beads and volume corrections on hydrodynamic parameters	84
5.4	Discussion	88

CHAPTER 6

PRODUCTION, PURIFICATION AND DERIVATISATION OF PNEUMOLYSIN 92

6.1	Introduction	92
6.1.1	Perfusion chromatography	94
6.2	Materials and methods	96
6.2.1	Bacterial stock Preparation	97
6.2.2	Production of crude toxin	97
6.2.3	Toxin purification and identification	97
6.2.4	Haemolytic assay	98
6.2.5	Derivatising pneumolysin with DTNB (Dithio(bis)nitrobenzoate)	99
6.2.6	Determining sample homogeneity	99
6.3	Results and discussion	100
6.3.1	Production and purification of pneumolysin	100
6.3.2	Efficiency of purification	105
6.3.3	Alternative strategies for optimising purification	107
6.3.4	Chemical derivatisation	111
6.4	Conclusion	118


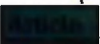


CHAPTER 7

AN ELECTROCHEMICAL METHOD TO STUDY THE LEAKAGE EVENTS IN PNEUMOLYSIN-INDUCED HAEMOLYSIS 119

7.1	Introduction	119
------------	---------------------	------------

7.1.1	Ion leakage from erythrocytes	119
7.1.2	The Enzymometer	120
7.2	Materials and methods	123
7.2.1	Sample preparation	123
7.2.2	Conductance measurements	124
7.2.3	Data analysis	124
7.2.4	Precautionary measures	124
7.3	Results	125
7.3.1	Conductance properties of constituent elements	125
7.3.2	Characteristics of toxin induced lysis in rabbit erythrocytes	130
7.3.3	The Effect of calcium and magnesium on toxin induced leakage	133
7.3.4	Pneumolysin dose response	137
7.4	Discussion	139
 CHAPTER 8		
CONCLUSION		141
 REFERENCES		143
 APPENDIX 1		161

TABLE OF FIGURES

- Figure 2.2.1 Illustrating the type of structural information obtained for an X shaped object (A) by crystallography (B) and from solution scattering (C). 9
- Figure 2.2.2 Constructive interference between two scattered waves, having wavelengths shorter than the molecule (which is represented as the shaded object). Since the scattering is 11
- Figure 3.1 Eight representations of a molecule (depicted by the shaded object in the middle) from which hydrodynamic parameters can be calculated. Whole body representations such as 21
- Figure 3.2 Three-dimensional models of (A) prolate and (B) oblate ellipsoids having two symmetrical axis, a and b , where a denotes the semimajor axis and b the semiminor axis..... 22
- Figure 3.3 An image of a molecule (A) rotated by a variety of different orientations to give an impression of how the scattering elements are distributed within that molecule. C and D..... 32
- Figure 4.1 The arrangement and dependencies of the four different types of libraries that constitute MOLEd are: , ,  and . The dependencies shown by 38
- Figure 4.2 Bead model displayed alongside sequence, chain and atom representation, on the 1D-3D interface..... 40
- Figure 4.3 An atomic representation of a short stretch of amino-acid residues is converted to bead models according to the *AtoB* (A) or *Trans* (I) method. In *AtoB* the atomic coordinates43
- Figure 4.4 All peptide bond beads except for proline residues are placed at the c.o.g. of the four main chain atoms shared between two residues. These are the C and O of the previous 44
- Figure 4.5 Bead models generated from atomic coordinates using the modified *AtoB* algorithm implemented in *SOMO*. The models generated at 3 Å, 5 Å, and 8 Å resolutions demonstrate how shape and bead numbers change at elevated resolutions. 45
- Figure 4.6 An ellipse, representing an elongated model, is used to demonstrate one of the flaws (identified by Byron, 1997) of isomorphous expansion as implemented by the *AtoB* 47
- Figure 4.7 Two overlapping spheres represented by differently shaded circles illustrate the effect of proportional reduction. The bead size of the smaller, darkly shaded sphere is 48
- Figure 4.8 Illustrates the removal of overlaps (iii) from a model containing overlapping beads (i). The consequence of overlap removal is a significant decrease in the volume of the model and a change in its surface area and topology (ii). 49

- Figure 4.9 The outer circle describes a hypothetical spherical molecule. In (1) 2 beads of the same diameter with a 50% overlap reside on the exterior of the molecule. With normal 52
- Figure 4.10 The difference between (i) translating a bead from the centre of gravity of the molecule (*) and (ii) translating it from a region closer to the site of translation (#). 53
- Figure 4.11 (i) The problem of translating a bead from a reference point close to the point of translation is illustrated here. A and B are two points on the grid representing regions close to 54
- Figure 4.12 Illustrates a mechanism for overlap removal that eliminates gaps which would have otherwise been present in the original outward translation procedure. In this picture, a 55
- Figure 4.13 Sequence of steps starting from the generation of a bead model from a short stretch of residues (i) using the *Trans* algorithm. Each bead is represented (ii) either as a 57
- Figure 4.14 Two spheres (i) represented as differently shaded circles; light (l) and dark (d) with a radius r_l and r_d and their centre points E and F , corresponding to the light and dark 58
- Figure 4.15 Example look-up table. The two values in the first line (highlighted in yellow) correspond to the number of lines in the look up table that is to be read in by *SOMO*, and the 63
- Figure 4.16 Chain and bead representation of a molecule (homology model for pneumolysin). Beads selected in the bead model are shaded in pink. They correspond with the pink residues on the chain model.68
- Figure 5.1 Extract from a PDB file of fibrinogen, fragment D (first box) and the *BeaMS* formatted files in the three columns of the second box. The first column block contains only 72
- Figure 5.2 Bead processing stages in both *AtoB* (B and C) and *Trans* (D to G) algorithms, illustrated for lysozyme. In *AtoB*, using a grid spacing of 5 Å, surface beads are placed first (in 75
- Figure 5.3 Atomic models for each of the four proteins: BPTI, RNase, lysozyme and CS are shown in the four columns of the first row. In the second row, each model was generated by 78
- Figure 5.4 Models for lysozyme generated by *Trans* using no hydration (this was simulated using a water volume of 0 Å), and then water volumes of 20 Å, 24 Å and 30 Å. Outward 80
- Figure 5.5 *AtoB* generated models for BPTI, at 1 Å resolution without outward translation (1) and at 5 Å resolution without (2) and with outward translation (3). 80
- Figure 6.1 Homology model of pneumolysin (Rossjohn *et al.*, 1998) visualised by VMD (Humphrey *et al.*, 1996) showing the single cysteine residue as spheres near the C-terminii.

Reduction of DTNB (pictured below the model) between the sulphur atoms causes it to split into two TNB molecules, which then bind to the cysteine residue. 93

Figure 6.2 POROS media is manufactured by adhering microspheres into particle clusters (1). Solutions pass between spaces found in the cluster giving molecules an increased 96

Figure 6.3 The first 280 minutes of growth for the *E. coli* strain M15 carrying the pkkwt plasmid. Growth was monitored by taking periodic spectroscopic measurements at 600 nm. 101

Figure 6.4 Fractions F1 to F8 from the HIC purification were run on 10% SDS PAGE (A). The gel is superimposed on the purification chromatogram monitored at 280 nm. 102

Figure 6.5 Pooled fractions (F5 to F8) from HIC were dialysed in PB(-S) and concentrated prior to injecting onto an AEC column. The chromatogram shows absorbance at 280 nm 104

Figure 6.6 Elution of contaminants from an AEC column at higher salt concentrations. Chromatogram shows absorbance at 280 nm (—) and 260 nm (—), and NaCl gradient () up to 1.5 M. 105

Figure 6.7 Toxin yield at different stages of the production and purification procedure. 107

Figure 6.8 A chromatogram of HIC purification of pneumolysin from crude extract is aligned with a 10% SDS PAGE gel showing protein bands from each fraction (F2 to F8), the 109

Figure 6.9 SEC chromatogram showing two peaks in absorbance at 280 nm. The first peak corresponds to the elution of pneumolysin aggregates whereas the second peak corresponds to monomeric pneumolysin. 113

Figure 6.10 Mass spectrometry traces for (A) pneumolysin, (B) pneumolysin-TNB conjugate, and (C) a 1:1 mixture of pneumolysin and pneumolysin-TNB. 116

Figure 6.11 SE studies for pneumolysin at 16 k (Δ) fitted with a dashed line and at 20 k (\square) fitted with a solid line, before (A) and after (B) SEC. 117

Figure 7.1 A simplistic circuit diagram for the enzymometer. Current flows into the reaction well from an AC source (V_i). The flow of current at the electrode in both directions 121

Figure 7.2 A frontal view of the enzymometer showing eight glass reaction chambers with substrate being pipetted into one of them. The unit is filled with water, which is continuously stirred and maintained at a fixed temperature. 122

Figure 7.3 Leakage rates were measured as the change in conductance over a period of time. Leakage rates for conditions (i) to (vii) were plotted on two separate graphs (A) and (B) to 127

Figure 7.4 Leakage and lytic events in rabbit erythrocytes monitored by conductimetry (▢) and absorbance (▼) measurements at 541 nm. Complete lysis occurred at $A_{541} = 0.64$

..... 132

Figure 7.5 Leakage response in both rabbit and sheep erythrocytes as a result of varying CaCl_2 concentration. A: 0.3% v/v rabbit erythrocytes with 1 μl pneumolysin (▲) and with

..... 134

Figure 7.6 Leakage rates taken from the gradients in Figure 7.5 were plotted against CaCl_2 concentrations for both rabbit (▲) and sheep (◆) erythrocytes. If leakage assumes a

135

Figure 7.7 Four slopes taken from graph B in Figure 7.5, at 0.1 mM (▼), 0.2 mM (◐), 1.0 mM (×) and 2 mM (※) CaCl_2 were used to illustrate calcium activation (A) and calcium

..... 136

Figure 7.8 A: 0.3% v/v rabbit erythrocytes with 10 μl pneumolysin. Conductance of the chamber was monitored at the following $[\text{Mg}^{++}]$: 0 mM (▲), 0.025 mM (◐), 0.05 mM (▼),

..... 136

Figure 7.9 A: 0.3% v/v sheep erythrocytes containing 0.5 mM CaCl_2 (▢) with 1 μl (▲), 2 μl (◐), 3 μl (▼), 4 μl (◐), 5 μl (◆), 6 μl (×) and 7 μl (※) of pneumolysin. B:

137

TABLE OF TABLES

Table 4.1 List of recent programs for generating bead models from PDB files (shaded) and neutron and X-ray scattering data. Not included in this list, because they have no program listing are permutations on the methods just described from (Banachowicz *et al.*, 2000; Perkins *et al.*, 1998) and (Zipper and Durchschlag, 1998). 41

Table 4.2 Overlap removal for both *AtoB* and *Trans* is a multi-step process requiring the reduction in the radius of a bead to remove overlaps. surface and buried correspond with beads either encountered at the surface or the interior of a molecule respectively. With the *Trans* algorithm, the location of the bead may have either an s.c. or p.b. extension denoting side chain and peptide bond respectively. Only those labels that are underlined are undergoing radius reduction. \Leftrightarrow refers to those bead types that are being compared with each other. 50

Table 4.3 *SwissPdbViewer* has recently changed the name of its program to *Deep View*. ♠ The source code was made available because *MOLEd* was being developed at the University of Glasgow, but this has not been made available to the public domain yet. ♥ Does not allow rotational manipulation of the model. ♣ Animations can be run through CHIME (Columbia Hypermedia Immersion Environment) (Dossick and Kaiser, 1999). 65

Table 5.1 Comparison between experimental and calculated values for the three test proteins: BPTI, RNase and lysozyme. In tables A and B, the comparison is as a function of water volume using *Trans* generated models with outward translation and popping set at 20% (only in table A) and no popping (only in table B). Table C compares models generated by *AtoB* at 1 Å and 5 Å resolutions with outward translation and at 5 Å resolution without 84

Table 5.2 Predicted parameters from models generated by *Trans*, using all the beads; just the exposed beads from the initial model and then from models that have been rechecked by *ASA*, with (c) and without (nc) summing the volume of the buried beads to the volume correction. All four different types of models were compared with experimental data (**Exp.**), references for which can be found in the previous table. 87

Table 5.3 Comparing Experimental (Exp.) data for CS (obtained from Zipper and Durchschlag (2000) with models generated by *Trans* (T^*) (70% popping) and *AtoB* (A^*) (5 Å resolution). 88

Table 6.1 Pneumolysin yield from a 1 L growth culture at different stages of expression and purification as determined by haemolytic activity measurements. 106

Table 7.1 List of chemicals and the volumes used to test the conductance properties for seven different solution compositions. Components were added in row order (top to bottom). Erythrocytes from human, horse, rabbit and sheep were used in condition (iv) and (v), while condition (vii) involved only rabbit erythrocytes. 126

Table 7.2 The data in tables A and B were obtained from Nelson (1967) (Nelson, 1967) for non-human erythrocytes and Farquhar (1962) for human erythrocytes. Table A shows a percentage distribution of the different types of lipids while table B contains the percentage composition of phospholipids. Some of the values obtained for rabbit red blood cells appear to be quite different from those obtained from Kazennov (1998) (Kazennov *et al.*, 1998). 129

ACKNOWLEDGEMENTS

There are many people that I would like to thank, including all my friends, flatmates, family and colleagues (particularly those in the South Lab) for their patience and good humour during my PhD. I would particularly like to mention those who have given up their time to help me with the more technical aspects of my project. Jeremy Beauchamp for guiding me through various bits of my project including crystallisation trials and size exclusion chromatography; Robert Gilbert for help and advice with pneumolysin; Paul Emsley, Stuart Mackay and Chris Edwards for providing me with network and operating system support; JC Nebel and Penny Dunabin for helpful discussions on certain aspects of the *SOMO* algorithm; Tim Mitchell and June Irvine for allowing me to use the BIOcad and other lab facilities, I'm sorry that I never got round to continuing the work with the pneumolysin mutants. Neil Issacs, Isobel Black and members of the crystallography group for allowing me to use lab and computing facilities. Jenny Search for her milligrams of help. Thanks Gordon for helping me out in 'sticky' situations. Aldo for helping me find my way around the lab at the ABC. Staff at RALS, ILL and Daresbury for assistance with using and understanding the instrumentation used for small angle scattering experiments. Andrew Dalke for help with modifying VMD.

Finally I owe a big debt of gratitude to my supervisors: Tony Lawrence for introducing me to the Enzymometer and for encouraging my interest in its application; Ron Poet for his help advice with programming and also for allowing me to use his *MOLed* libraries from which I could build *SOMO*; Mattia Rocco for fine wine, enlightened discussions and whose ideas and careful guidance have ultimately shaped the *SOMO* program. I also thank him for being very open to my questioning regardless of its simplicity and often stupidity, for proof reading several of my chapters and for his support at all times. Finally to Olwyn Byron, whose overall supervision of my project and support to let me explore fields that I found interesting (albeit justifiably reluctant on some occasions) has made this project for me, both rich and fulfilling. I am also grateful to her for stepping in on numerous occasions to sort out problems experimental, mathematical and grammatical, and also for making this thesis more legible than it would have otherwise been. I thank you all for your enthusiasm, wit and kindness.

This project was funded by a BBSRC special studentship and an EMBO Short Term Fellowship that supported a collaboration with M. Rocco, at the Advanced Biotechnology Centre (ABC), Genova, Italy. I also acknowledge travel bursaries from both the University of Glasgow and the ABC.

I would like to dedicate this thesis to my parents, my sister Neema, and my grand mother, Appucho.

 ABBREVIATIONS AND SYMBOLS

<i>1D-3D</i>	Computer program for displaying atomic models
2θ	Scattering angle
AEC	Anion exchange chromatography
<i>ASA</i>	Program for calculating Accessible Surface Area.
<i>AtoB</i>	Atoms to Beads – computer program for generating bead models
AUC	Analytical ultracentrifugation
B_i	Atomic scattering length for a neutron at point i in a molecule
B	Second virial coefficient (ml mol g^{-2})
B	$3N \times 3N$ supermatrix containing the $N \times N$ matrix, Q_{ij} .
Ξ	6×6 friction tensor, which is the inverse of B
<i>BeaMS</i>	Bead Modelling System – complete software package for bead modelling.
BPTI	Bovine pancreatic trypsin inhibitor
C	Velocity of electromagnetic radiation (ms^{-1})
C	Concentration of sample (ml g^{-1})
c.o.g.	Centre of gravity
cryo-EM	Cryo-electron microscopy
CS	Citrate synthase
CV	Column volume (ml)
D	Diffusion coefficient ($\text{cm}^2 \text{s}^{-1}$)
D_2O	Heavy water
D_{max}	The maximum dimension of a macromolecule (nm)
DTNB	Dithio(bis)nitrobenzoate
ϵ	Extinction coefficient ($\text{cm}^2 \text{mg}^{-1}$)
F_i	Atomic scattering length of electrons at point i in a molecule
ζ	Frictional coefficient (gs^{-1})
FPLC	Fast Protein Liquid Chromatography
h	Planck constant ($6.626 \times 10^{-34} \text{ Js}$)
HIC	Hydrophobic interaction chromatography
HRTs	High resolution techniques
I	Scattering intensity
$I(0)$	Forward scattering
IPTG	Isopropyl thio-galactoside
k	Boltzmann's constant
δ	Hydration (or solvation) of a particle ($\text{g H}_2\text{O g}^{-1}$)
λ	Wavelength of electromagnetic radiation (nm)

MOPS	3-(N-morpholino) propanesulphonic acid
M_r	Molecular mass (g mol^{-1} or Da)
N	Number of beads in the model
N_A	Avogadro's number
p	The axial ratio for an ellipsoid
p.b.	Peptide bond bead
PBS	Phosphate buffer saline
PB(-S)	PBS preparation but without the saline component
0.7 M PBS	PBS preparation with 0.7M NaCl
PC	Phosphatidylcholine
PDB	Protein data bank
PE	Phosphatidylethanolamine
PROMOLP	Computer program for calculating biophysical parameters from amino acid sequence data
PS	Phosphatidylserine
ω	Rotational velocity (rad s^{-1})
OO	Object oriented
Q	Scattering vector
Q_{ij}	Supermatrix corresponding to the hydrodynamic forces acting on a molecule.
$[\eta]$	Intrinsic viscosity ($\text{cm}^3 \text{g}^{-1}$)
η_0	Solvent viscosity (Poise)
σ	Radius of a sphere (nm or Å)
ρ	Density of solvent (g ml^{-1})
$^s\rho$	Scattering density for a group of atoms
R	Distance vector
r_{bnd}	Distance between the molecule and the centre of a spinning centrifuge rotor (cm)
R_{ij}	The separation between two scattering elements i and j in a molecule (nm or Å).
R_g	Radius of gyration (nm or Å)
<i>Rasmol</i>	Molecular visualisation program
RAZ	Extension program for <i>Rasmol</i> to visualise bead models
Rnase	Ribonuclease A
S	Sedimentation coefficient (S); 1 S is equivalent to 1×10^{-13} s
SANS	Small-angle neutron scattering
SAS	Small-angle scattering

SAXS	Small-angle X-ray scattering
s.c.	Side chain bead
SDS-PAGE	Sodium-dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
<i>SOMO</i>	Solution Modeller
SM	Sphingomyelin
STs	Solution techniques
<i>SUPC</i>	Computer program for calculating hydrodynamic parameters
<i>T</i>	Hydrodynamic interaction tensor
T	Absolute temperature (K)
<i>Trans</i>	Computer program for <i>Transforming</i> atomic models to bead models
<i>V</i>	Volume (nm ³ or Å ³)
<i>v</i>	Velocity (ms ⁻¹)
\bar{v}	Partial specific volume (ml g ⁻¹)
<i>VMD</i>	Visual Molecular Dynamics, a molecular visualisation program

CHAPTER 1

INTRODUCTION

1.1 CHARACTERISING PROTEIN STRUCTURE AND INTERACTIONS

In the everyday world, objects can be physically described using the full range of human perceptions. While at the molecular level, the physical description of structures is limited to quantities that relate to their size, shape and dynamic properties. Properties such as mass (M), volume (V) and radius of gyration (R_g)¹ are general descriptions that are commonly determined by hydrodynamic (e.g. analytical ultracentrifugation – AUC (Laue and Stafford, 1999)) and solution scattering techniques (e.g. small-angle scattering - SAS² (Perkins, 1988a)), collectively referred to as solution techniques (STs). Analysis by these techniques is performed while the molecules are in solution in contrast with techniques such as X-ray crystallography and electron microscopy both of which require the protein environment to be altered either by crystallising the sample or by drying it. The consequence of this preparation of the sample is that it may inadvertently alter the final conformation of the protein, either from physical handling of the sample or from differences in the physiochemical conditions of the sample environment. While conformational changes due to experimental preparation is less likely to occur with STs, it does not imply that such changes do not occur during analysis. Take SAXS for instance; X-rays can destabilise proteins in solution (Weik *et al.*, 2000) and also cause them to aggregate (Gregory *et al.*, 1987). The sensitivity of these proteins to radiation damage is dependent on the mass of the molecule (and not its molecular volume or shape) (Osborne *et al.*, 2000).

The hydrodynamic and scattering information derived from the above mentioned STs tends to be of low resolution and lacks the level of detail offered by high resolution techniques (HRTs) such as X-ray crystallography and NMR, and medium resolution techniques such as cryo-electron microscopy (cryo-EM). HRTs can resolve protein structures to their atomic level and in doing so set the standard in terms of the level of detail achievable for molecular structure determination. But HRTs are limited by a number of factors. For instance, they require large quantities (milligrams) of very pure protein. With X-ray crystallography, the crystallisation of some proteins can be problematic, particularly if that protein happens to be either very large, quite flexible, or insoluble in polar solvents. The screening of different physiochemical conditions for crystal growth can also be very time consuming. And, like

¹ The R_g is essentially a measure of macromolecular elongation based on the distribution of matter within the macromolecule.

² SAS in this thesis refers to scattering of either neutrons (SANS) or X-rays (SAXS).

SAXS, the damaging effects of X-ray radiation (even under cryogenic conditions) can effect the stability of the molecule, with reports of cleaved disulphide bridges along with a general loss of definition, particularly with acidic residues (Weik *et al.*, 2000)³.

Even if a molecule is successfully crystallised, satisfactory diffraction from the crystal is not assured. Molecules can exhibit a detrimental degree of flexibility within the crystal lattice. The degree of flexibility is indicated by the B factor or temperature factor. This factor represents the mean displacement of the atom's position and hence is an indication of regions of flexibility. NMR on the other hand is a solution technique which generates a family of structures. These structures and hence the molecular flexibility can be seen by displaying backbone representations of the model (Barlow *et al.*, 1993). Superposition of these structures will give an average conformation which is used to display the atomic model. NMR does not suffer the same setbacks as crystallography but is rather limited by its inability to resolve structures for molecules of mass greater than 30 kDa.

Cryo-EM has recently emerged as a technique that can solve structures to resolutions of up to 7.5 Å as reported by Matadeen *et al.* (1999) for the large ribosomal subunit of *Escherichia coli*. It is particularly useful for solving the structure of large molecular complexes that are practically impossible by NMR and very difficult by crystallography, particularly if the components of the complex interact weakly. However if individual components of a large complex can be solved separately by HRTs, it should then be possible to link these subset structures using the cryo-EM model of the entire complex as a template (Mueller *et al.*, 2000). This fitting can be facilitated by heavy atom labelling, site directed mutagenesis and other methods listed in Tao and Zhang (2000). Fitting has also been automated using neural networks to fit actin monomers on to an electron micrograph of the actin filament (Wriggers *et al.*, 1998). The concern is that the components of the complex remain stable i.e. not flexible or does not alter its conformation when separated from the complex (Betts and Sternberg, 1999).

More recently there has been the prolific use of computational methods for predicting protein structures (Webster, 2000). This has partly come about by the rapid growth in protein sequence data particularly as the result of efforts by those institutions involved in genome sequencing (Burley, 2000). Two basic approaches for the utilisation of sequence information for structure prediction are (i) comparative modelling and (ii) *ab initio* prediction (Jones, 2000). The first method relies on comparing the sequence of an unknown structure with a database of sequences of known structures while the second relies on rules governing how the primary sequence should fold into its tertiary structure. The first approach is particularly useful if the

³ These effects may be due to the radiation or a consequence of free radicals produced from its interaction with the solvent environment.

protein sequence of the unknown structure has greater than 30% sequence homology to structures that have already been solved experimentally. However, if the structures of the homologous sequences have not been solved, or if the protein of interest does not have any homologous sequences, as is the case with ORFans (Open Reading Frames that show no similarity to proteins of other genomes⁴) (Fischer and Eisenberg, 1999), then it is necessary to use the second approach.

A variety of computer programs is currently available for protein structure prediction (Webster, 2000). The accuracy and capability of these programs are assessed every two years by the CASP exercise (Critical Assessment of techniques for protein Structure Prediction⁵ (Mosimann *et al.*, 1995)) with results published at <http://predictioncenter.llnl.gov> indicating their varying degrees of success. Computer programs are also available for predicting whether and how two structures interact to form a complex. For example, if structural information is available, then popular strategies such as surface electrostatics and shape complementarity (Gabb *et al.*, 1997), or shape, hydrophobicity and solvent accessibility (Jones and Thornton, 1997) are commonly employed. If only sequence information is available then programs such as SVM (Support Vector Machine) (Bock and Gough, 2000) can predict those proteins that are likely from its sequence. The program uses a neural network that has been trained on the protein interaction database (DIP) (Xenarios *et al.*, 2000). The availability and accessibility of sequence information along with rapidity and low cost in generating atomic resolution structures has made structure prediction a useful first step at the early stages of structure function studies.

A drawback with predictive methods particularly when they involve predicting the conformation of large multisubunit complexes lies in the evaluation of the success of the prediction. This can be done by comparing the predicted structure with a structure that has been solved experimentally by HRTs. But the popularity of predictive methods derives from the inherent difficulties associated with structure determination by HRTs. Even if the structure of a molecular complex was solved and validated⁶ (Laskowski *et al.*, 1998), it does not necessarily imply that the complex adopts the same structural conformation in its native solution environment. While there are numerous examples of protein quaternary structures determined by STs differing from those determined by crystallography (Svergun *et al.*, 1997; Svergun *et al.*, 2000), there is a growing body of evidence that suggests proteins to be much

⁴ Using the complete genome sequence from a selection of bacteria, Fischer and Eisenberg (1999) showed that for some bacteria, up to 25% of the genes had no similarity with known sequences.

⁵ Their main measure for model quality is the RMS (root mean square) deviation between the predicted model and its real structure.

⁶ Validation can take the form of experimental validation where the R_{free} ratio and the Q-factor of a protein are indicators of the reliability of data from crystallography and NMR respectively and structural validation, which uses computer programs that for the most part compare the protein's geometry with standard values derived from datasets of high resolution protein structures (Laskowski *et al.*, 1998).

more disordered than depicted by high resolution crystallographic structures (Dunker *et al.*, 2001). The implication is that crystallographic structures may simply represent any one of a possible number of states (ordered; globular and random coil) and conformations of a molecule. Capturing this disorder or uncertainty as a three dimensional model has been attempted with computer programs such as *PROTEAND* (Altman *et al.*, 1995), whereas bead modelling programs such as *AtoB* (Byron, 1997) have been used to monitor more clearly conformation changes (Errington *et al.*, 1999).

A bead model within the context of this thesis is a multisphere representation of a single protein or a protein complex. The arrangement of these spheres (or beads) is governed by a set of rules that allows such models to be interpreted in terms of their hydrodynamic or solution scattering parameters. These parameters include M , V , R_g , the frictional coefficient, ζ , and the diffusion coefficient, D (the latter two parameters are derived from observations of molecules dynamic behaviour which is then used to infer shape information). Bead models that are used in the calculation of hydrodynamic parameters are commonly referred to as hydrodynamic bead models, whereas bead models that are used in the calculation of solution scattering parameters are known as Debye sphere models. The rules governing the arrangement of beads in both types of models are different but they are also flexible enough to allow both types of parameters to be calculated from a single model. Generating models that are a compromise between Debye and hydrodynamic models may reduce the accuracy of the calculated parameters, but has the advantage of facilitating the simulation of solution parameters generated by both types of STs.

One procedure for evaluating a predicted structure is by constructing a bead model that represents its shape. From this its solution parameters can be compared with the same parameters determined experimentally. If the conditions for the experimental parameters have been standardised⁷ and both experimental and theoretical parameters are relatively close to each other, then the model is considered to be plausible. But this does not confirm the solution structure of a molecule purely because variations on a single shape (or conformation) can also lead to identical theoretical parameters. Therefore bead modelling in this sense is better used as method that will filter out implausible structures. Obviously confidence in structural prediction using STs will improve with the application of more filters. This is one of the reasons why there is an interest in combining hydrodynamic and Debye sphere modelling strategies.

In summary, this section has highlighted some of the advantages and disadvantages of a variety of popular experimental and computational techniques for determining the structure

⁷ For hydrodynamic techniques, this involves measuring at 20 °C and correcting for infinite dilution using water as its solvent.

and interaction between macromolecules. Each technique provides subsets of information that are used in characterising the whole system. A major part of this thesis focuses on the further development of two bead modelling programs: *AtoB* (Byron, 1997) and *Trans*, (which belongs to the *BeaMS* suite of programs (Spotorno *et al.*, 1997)). Both represent two different methods for bead model generation from atomic resolution data; with *Trans* focussing more towards modelling hydrodynamic data whereas *AtoB* is more applicable for modelling SAS data. Modifications of the algorithms used in both programs have been implemented in the computer program *SOMO* (Solution MOdeller), the aim of which is to test each algorithm with hydrodynamic and scattering data in order to develop a single algorithm that is best utilised with both methods. Its development and testing as described by the first part of this thesis is part of a much larger project that aims to characterise the interaction between proteins from different experimental systems, including that between pneumolysin and a panel of anti-pneumolysin monoclonal antibodies (de los Toyos *et al.*, 1996).

Pneumolysin, a 53 kDa membrane damaging thiol-activated toxin expressed by the gram positive bacterium, *Streptococcus pneumoniae* is one of the most important causes of community acquired pneumonia (CAP) and is also responsible for otitis media, sepsis and meningitis. Experimental studies have shown the route of infection to begin with colonisation of the nasopharynx and then spreading either upwards through the Eustachian tube and into the middle ear to cause otitis-media or downwards into the alveoli of the lower respiratory tract to cause lobar pneumonia (Cundell *et al.*, 1995). The initial response to a pneumococcal infection is the activation of the host immune defences. This leads to an inflammatory response, culminating in the eventual lysis of the pneumococci. Lysis of the cell releases the pneumolysin toxin, in its monomeric form. These monomers are thought to integrate into the membranes of eukaryotic cells containing cholesterol. Once in the membrane, the monomers are thought to self-associate with other pneumolysin monomers by a process of lateral diffusion, forming large oligomeric ring and arc shaped structures that have been confirmed by electron microscopy (Morgan *et al.*, 1994).

Pneumolysin is 471 residues in length and belongs to the thiol activated family of cytolytins which include perfringolysin and streptolysin both of which are produced by *Clostridium perfringens* and *Streptococcus pyogenes* respectively. This family has the characteristic signature sequence 'E-C-T-G-L-x-W-E-W-W' located towards the C-termini (between residues 427 and 436 in pneumolysin). All members of this family possess a single cysteine residue which is essential for membrane binding (Iwamoto *et al.*, 1987). The membrane binding domain for pneumolysin toxin has been identified through a series of C-terminus deletions. Studies using site directed mutagenesis of the C-termini region of pneumolysin suggests that the final 12 C-terminal amino-acid sequence is important for the binding of the toxin to the cell membrane. A comparison of the C-termini sequences of five thiol-activated toxins showed proline to be

conserved in those sequences. Proline, found at position 462 in pneumolysin, when substituted with serine was shown to give a 27% reduction in haemolytic activity and a 90% reduction in cell binding activity (Owen *et al.*, 1994).

A variety of monoclonal antibodies raised against the pneumolysin toxin have been used to further understand its functional role. De los Toyos *et al.* (1996) describes the results of the interaction of several IgG monoclonal antibodies with the pneumolysin toxin. They showed two routes for blocking the cytolytic action of the toxin: blocking the toxin from integrating into the membrane or by preventing the monomeric units of the toxin from oligomerising. Understanding how these antibodies (or those mutations performed by Owen *et al.* (1994)) affect toxin activity can be explained from a structural perspective by building computer models from experimental data and by means of biochemical assays that monitor its reaction kinetics. Ensuring that the computer models are an accurate representation of the experimental data is imperative for successful modelling as is a sensitive method for monitoring toxin activity. Hence, a considerable amount of effort was devoted to improvements to the modelling procedure and towards designing a protocol for measuring haemolytic activity.

1.2 THE PRESENT STUDY

There were three basic aims to this project: (1) the development of a computer program for generating bead models (*SOMO*); (2) the development of a protocol for purifying pneumolysin and preparing it for further biophysical characterisation; and (3) designing a protocol to monitor the pore forming activity of the toxin. Most of the methods described in this project were used previously by Gilbert *et al.* (1998) to characterise the pore forming mechanism of the toxin. The difference between those studies and the one described in this project is that a greater emphasis has been placed on the development of some of the techniques used in those studies, particularly the computer program *SOMO*.

The bead models used by Gilbert *et al.* were generated using *AtoB*. Although this program can generate bead models for pneumolysin and other proteins, the representation of the hydration shell around the model was found to be inadequate, particularly for elongated models (Byron, 1997). Carrasco *et al.* (1999) had also noted that models with overlapping beads (as was present in the *AtoB* models) can give erratic results, because the theoretical basis for the hydrodynamic predictions are not strictly valid, as opposed to models with no overlapping beads. These two issues as well as an interest in improving the actual interface for visualising the models were the initial factors that shaped *SOMO*'s development.

SOMO was written in C++ using a recommended procedure for developing a set of object libraries (Huber and McCammon, 1999). The majority of these libraries belonged to the program *1D-3D* (Belhadjmostefa *et al.*, 1991) which is currently being modified as *MOLed* (*MOLecular editor*). In order to appreciate how solution parameters can be calculated from bead models, it is beneficial to provide some background information on the two types of STs that have been the focus of this thesis: SAS and AUC. The actual principles underlying both techniques are described in the following chapter, with chapter 3 providing an overview of the theoretical basis for hydrodynamic and SAS calculations from bead models. *SOMO* uses modifications of the algorithms used by the *AtoB* and *Trans* computer programs for generating bead models. Both algorithms as well as the development of *SOMO* and a discussion of its features are given in chapter 4. Chapter 5 tests both algorithms by comparing how the hydrodynamic predictions made by *SOMO* compare with experimental values for four test proteins that have been previously well characterised experimentally by others.

Pneumolysin was extracted from recombinant *Escherichia coli* containing a copy of the pneumolysin gene (Mitchell *et al.*, 1989). The toxin was purified from the bacterial cell extract by perfusion chromatography (Sheehan, 2000) as described in Chapter 6. This approach was found to be faster than the traditional FPLC (fast protein liquid chromatography) methods previously used for purifying pneumolysin (Mitchell *et al.*, 1989). Previous studies on pneumolysin have included structural and kinetic analysis of its pore forming capabilities using SANS, AUC, cryo-electron microscopy and haemolytic assays (Gilbert, 1998). This present thesis includes preliminary work in the development and testing of an electrochemical method for monitoring the pore forming activity of pneumolysin. The method which is described in chapter 7 was used to monitor the activity of toxin under different ionic conditions and contrary to previous findings, showed that the leakage of ions from the cell occurred prior to complete cell lysis.

CHAPTER 2

EXPERIMENTAL DETERMINATION OF SOLUTION PARAMETERS

2.1 SMALL ANGLE SCATTERING

The underlying principles of Small Angle Scattering (SAS) are not too dissimilar from those which form the basis for X-ray crystallography. Both require the analysis of radiation diffracted by atoms with wavelengths, λ ranging from $100 \text{ \AA} - 0.1 \text{ \AA}$, which are short enough to resolve a covalent bond (about 1.2 \AA). The main difference is that with crystallography matter is ordered in a crystal lattice, thus holding the protein conformation in a semi-rigid state. So, incident radiation scattered by this lattice can then be resolved into a three-dimensional model (Figure 2.2.1, A and B). Molecules in solution are however randomly oriented, so the information obtained represents an average of all their different orientations (Figure 2.2.1 C).

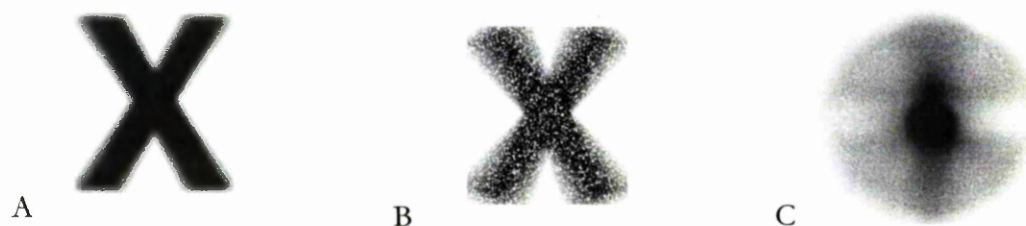


Figure 2.2.1 Illustrating the type of structural information obtained for an X shaped object (A) by crystallography (B) and from solution scattering (C).

High flux radiation can be generated from either synchrotron or reactor sources. Examples of X-ray and neutron synchrotron sources include the SRS facility at Daresbury, UK and at ISIS, Rutherford Appleton Laboratory in Didcot, UK, respectively. An example of a reactor source that generates neutrons is the high flux reactor source at the Institut Laue Langevin, Grenoble, France. This section, focuses on the principle of SAS as used to determine general shape information for biological molecules.

2.1.1 BASIC PRINCIPLES

Incident radiation in the form of X-rays or neutrons is scattered when it interacts with the electron cloud of an atom or its nucleus respectively. Scattering can either be elastic or inelastic, the main difference being that no energy⁸ is lost in the elastic scattering process. Inelastic scattering of neutrons can be used to study the motion of atoms in molecules (at least when neutrons are used⁹), whereas elastic scattering is used in SAS measurements to study the overall structure of the molecule. Elastic scattering with neutrons can be of two types coherent and incoherent. The difference is dependent on the spin of the nuclei. Nuclei with zero spin such as ¹²C and ¹⁶O do not scatter incoherently. Scattering can also occur from weak magnetic interactions with the electron cloud of an atom. The effect of incoherent scattering and scattering by the electron cloud is insignificant except in the case of hydrogen, which has a large negative neutron scattering length.

The magnitude of the scattering length of a slow neutron¹⁰ or X-ray is unique for each element (and for each isotope in the case of neutrons). The X-ray scattering length unlike that for neutrons, increases with atomic number. So in the case of X-rays it is difficult to distinguish the scattering contributions made by elements of similar mass. The neutron scattering length of nuclei (except those of hydrogen and deuterium) is smaller than that for X-rays and shows no correlation with the atomic number of the elements. The neutron scattering length for hydrogen (¹H) and deuterium (²H) is -3.742 fm and 6.671 fm respectively. The X-ray scattering length for both elements is 2.81 fm.

Scattering measurements rely on the contrast between the scattering caused by elements in the molecule and those in solution. Different types of biological molecules or their components e.g. lipids, proteins, carbohydrates and nucleic acids all have different scattering length densities. With neutrons, their scattering length densities lie between the scattering length density of deuterated and un-deuterated water. If the macromolecule is a multi-component system, then each component can be identified by matching the scattering length density of the partially deuterated buffer with that of the component. This involves varying the deuterium concentration of the buffer to identify match points for each of the components; a technique

⁸ The energy (E) for any form of electromagnetic radiation is determined by $E=hc/\lambda$, where h is the Planck constant (6.626×10^{-34}), c the velocity of the radiation and λ its wavelength. For neutrons, $E=h^2/2m\lambda^2$, where m is the mass of the neutron (1.675×10^{-24}). The energy of a neutron is approximately four orders of magnitude less than that of an X-ray of the same wavelength.

⁹ The energy difference between the incident and inelastically scattered neutron is much greater than for X-rays (Blume, 1975). For neutrons, the difference is dependent on the motion of the particle and so specific modes of motion can be observed by measuring inelastic scatter at specific angles. For further detail see Egelstaff (1975).

¹⁰ Neutrons possess a range of energy distributions. Neutrons with the lowest energy distribution (~ 0.05 eV) are referred to as thermal or slow neutrons. Fast neutrons possess the highest energy, greater than 1 MeV on fission. Any future reference to neutrons in this thesis will assume that they are slow neutrons unless stated otherwise.

known as contrast variation. Therefore a non-deuterated sample placed in a deuterated environment would give contrasting scattering length densities and is the basis for a variety of biological SANS experiments.

Coherent scattering of both neutrons and X-rays is determined by Bragg's law, $\lambda = 2R_{ij} \sin \theta$, where λ is the wavelength of either the X-ray or neutron; R_{ij} the separation between two scattering elements i and j in a molecule, and 2θ the angle of diffraction (Figure 2.2.2).

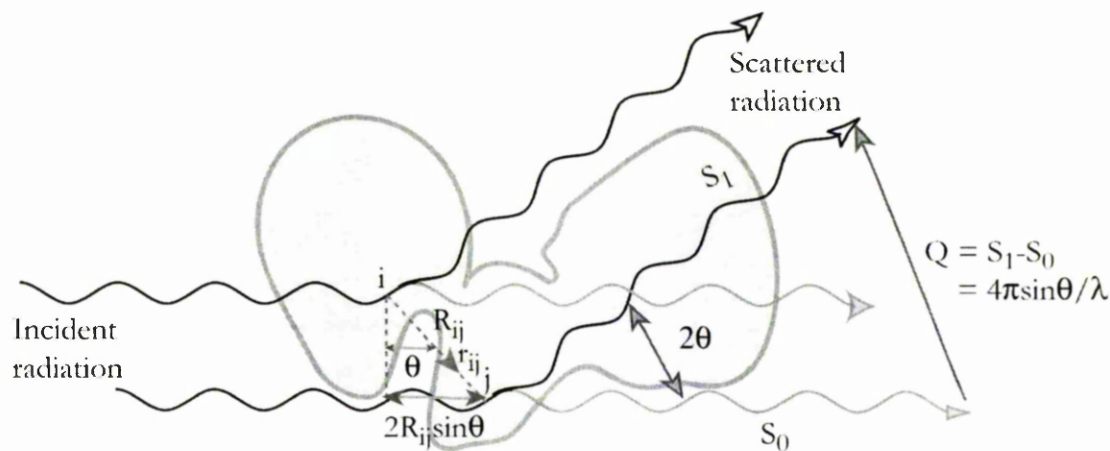


Figure 2.2.2 Constructive interference between two scattered waves, having wavelengths shorter than the molecule (which is represented as the shaded object). Since the scattering is elastic, the amplitudes of the incident and the scattered radiation are the same. In crystallography, the angle θ is used to determine the distance relationship between different scattering elements. In this example, the distance between two scattering elements i and j is R_{ij} , whereas r_{ij} is the vector that connects i to j . For SAS, the scattering angle (2θ) is incorporated into the scattering vector Q . The scattering intensity (I) is measured as a function of Q to determine the overall size and shape of the macromolecule. While measurements taken at wider angles (which is not the concern of this thesis) can be used to probe the molecular structure in more detail.

The intensity of scattered radiation $I(Q)$ is measured as a function of the scattering vector, Q , where $I(Q)$ is given by the Debye equation as

$$\overline{F^2(Q)} = \overline{I(Q)} = \sum_i \sum_j f_i f_j \frac{\sin(r_{ij} Q)}{r_{ij} Q} \quad \text{Equation 2.1}$$

and Q

$$Q = 4\pi \frac{\sin \theta}{\lambda} \quad \text{Equation 2.2}$$

with F the molecular structure factor, f_i and f_j (or b_i and b_j) the atomic scattering lengths for electrons (or nuclei) at points i and j in the molecule, connected by the vector r which is $r_{ij} = |r_i - r_j|$, where r_i and r_j are their respective coordinates. Because the molecules are in solution, they are oriented randomly relative to the incident beam. So $I(Q)$ is an average value taken over all possible orientations (indicated by the superscript bar) and is only dependent on the magnitude of Q .

The Debye equation assumes that all the molecules are identical, monodisperse, randomly oriented and do not interact with each other. It also assumes that the molecules are *in vacuo*. The fact that they are not can be accounted for by subtracting the scattering caused by the solvent from that caused by the molecule. Now because the structural resolution for solution scattering experiments is at best between 1 – 4 nm (Perkins, 1988a) i.e. beyond inter atomic distances¹¹, the scattering caused by both molecule and solvent can be expressed in terms of its scattering length density. This requires grouping atoms so that the scattering density for each group is given as ${}^s\rho = \sum f_i / V$ (f can be replaced by b in the case of neutrons), where V is the volume of that group. Applying this modification to equation 2.1 gives:

$$\overline{I(Q)} = \int_V \int_V \rho(r_i) \rho(r_j) \frac{\sin(rQ)}{rQ} d^3 r_i d^3 r_j \quad \text{Equation 2.3}$$

where $\rho(r_i)$ and $\rho(r_j)$ denote scattering elements composed of a collection of atoms. The contrast between the scattering caused by a molecule to that in solution is given by the difference in their scattering densities:

$$\Delta {}^s\rho = {}^s\rho_v - {}^s\rho_s \quad \text{Equation 2.4}$$

Where ${}^s\rho_v$ and ${}^s\rho_s$ are the scattering length densities (scattering length per unit volume) of the molecule and solvent respectively. The combination of equation 2.3 and equation 2.4 gives the amendment that was required to account for the presence of solvent molecules in solution:

$$\overline{I(Q)} = \int_V \int_V ({}^s\rho(r_i) - {}^s\rho_s) ({}^s\rho(r_j) - {}^s\rho_s) \frac{\sin(rQ)}{rQ} d^3 r_i d^3 r_j \quad \text{Equation 2.5}$$

Because scattering is symmetrical around the incident beam, the captured data can be circularly averaged to give a plot of $I(Q)$ as a function of Q .

2.1.2 DERIVING PHYSICAL PARAMETERS FROM SMALL ANGLE SCATTERING

Molecular mass (M_p , g/mol), radius of gyration (R_g , nm or Å) and molecular volume (V , nm³ or Å³) are just some of the physical parameters that can be measured using small-angle scattering. In the former case, the intensity of scattering when 2θ is zero, $I(0)$ is directly proportional to the mass of the molecule (cited in Perkins, 1988b). This is because at this angle, the scattered waves are in phase and scattering is at a maximum, directly reflecting the sum of all scatterers. A value for $I(0)$ and R_g can be determined at small Q values using the Guinier approximation¹²:

$$\ln I(Q) = \ln I(0) - R_g^2 Q^2 / 3 \quad \text{Equation 2.6}$$

where a plot of $I(Q)$ against Q^2 gives the R_g which is proportional to the gradient of a linear slope that intercepts the $I(Q)$ axis at $I(0)$. $I(0)/c$ is proportional to M_p and so the absolute molecular weight of a molecule can be determined using a sample of known molecular weight, concentration and having a similar Guinier Q range (Perkins, 1988b). Kratky (1963) had originally proposed using the intensity of scattered X-rays to determine M_p ; and while this approach can also be applied to neutrons, Jacrot and Zaccai (1981) used the incoherent scattering from water to determine M_p . Their method relied on the relatively strong incoherent scattering of neutrons from water, which for considerably dilute solutions would come mostly from the solvent. Because incoherent scattering is angle independent, it is unaffected by the distribution of scattering centres in the molecule. By taking the incoherent scattering of water in the forward direction ($I_{inc}(0)$), the expression given by Jacrot and Zaccai (1981) for $I(0)$ is

$$\frac{I(0)}{I_{inc}(0)} = A_f \frac{4\pi T_s}{(1 - T_{H_2O})} M_p N_A \times 10^{-3} \left(\frac{1}{M_r} (\sum b^{-s} \rho_s V) \right)^2 \quad \text{Equation 2.7}$$

¹² The resolution of a diffraction pattern is given by $2\pi/Q_{max}$, which can be derived by substituting λ (in equation 2.2) with the Bragg equation i.e. $Q=4\pi\sin\theta/2R_{ij}\sin\theta$, so that the shortest observable atomic separation, R_{ij} is attained when Q is at its maximum (Q_{max}).

where T_s and T_{H_2O} are the measured transmissions of the sample and water respectively, N_A is Avogadro's number, t is the pathlength of the sample (cm). For water with $t = 1$ mm, which is irradiated by neutrons ($\lambda > 1$ nm), the incoherent scattering is uniform in all directions i.e. isotropic. While for shorter wavelengths, the recoil of a nuclei from collisions with neutrons results in strong incoherent scattering in the forward direction, which has an angular dependence that is most pronounced at the shorter wavelength. This is corrected for by the anisotropy factor, A_f . The wavelength of neutrons produced on LOQ (the SANS station at RAL) is between 0.2 – 1.0 nm. This is too small for using the Jacrot and Zaccai (1981) method, therefore the comparative approach using molecules of known M_r is used instead (Kratky, 1963; Perkins, 1988b).

Having determined a value for M_r , it is possible to use this value to determine the molecular volume. Alternatively, volume determination can be made directly from $I(0)$ by matching out the scattering from the molecule using deuterated solvent (Perkins, 1986).

$$I(0) = (\sum_i b_i - s \rho_s V)^2 \mathcal{A} = \Delta^s \rho^2 V^2 \quad \text{Equation 2.8}$$

\mathcal{A} is a factor which includes the incident beam intensity (number of neutrons/cm²s), the sample thickness t , concentration C , its transmission T_s and more generally all the geometrical parameters of the experiment. So a plot of $\sqrt{I(0)/CtT_s}$ against $s\rho_s$ is linear from which V can be determined. For neutrons, V relates to the anhydrous volume, while for X-rays (by replacing the neutron scattering length (b) in equation 2.8 with f , for X-rays) the volume includes the hydration shell around the molecule.

The M_r of a molecule can also be determined with the AUC, along with both s and D to give ζ and V ($M_r + \bar{v}$). ζ is commonly used as a measure of particle elongation by comparing its shape with that of a sphere of equivalent volume (ζ_0). The ζ/ζ_0 ratio is in many ways similar to an R_g/R_0 ratio, where R_0 is the radius of gyration of the sphere. While both ratios describe particle elongation, they share a non-linear qualitative correlation which means that one parameter cannot be used to determine another Perkins (1988).

¹² Guinier analysis is only valid when the Q values lie between a minimum value where $QD_{\max} \leq \pi$, (D_{\max} is the maximum dimension of the macromolecule) to a maximum value that depends on the shape of the particle (e.g. QR_g should be less than 1.3 or 0.7 for sphere or rod shaped molecules, respectively). It is also sensitive to aggregates, which cause the $\ln I(Q) \propto Q^2$ plot to curve upwards. Objects approaching the size of dust particles are however too large to be detected by small angle X-ray or neutron scattering, but are often problematic in light scattering measurements, particularly with multi-angle laser light scattering, a technique commonly used for measuring the R_g of larger macromolecules.

2.2 ANALYTICAL ULTRACENTRIFUGATION

The analytical ultracentrifuge (AUC) is a high speed centrifuge that uses optical detection (either via absorbance or refraction) to measure the concentration (C) of a sample as a function of its radial position (r) in the sample holder. The sample holder consists of a centrepiece sandwiched between two quartz or sapphire windows. In its simplest form the centrepiece has a pair of hollowed sector shaped channels that are symmetrically aligned, with the widest part of the sector being furthest from the rotor centre. The shape of each channel follows imaginary radial lines emanating from the rotor centre, so that sedimenting molecules have a reduced likelihood of colliding against the side walls when the rotor is set spinning. This sector shape unfortunately complicates the mathematical analysis of the experiment, but a partial solution has been found in the form of the Lamm equations (Cantor and Schimmel, 1980). The construction of the sample holder is such that it is able to withstand high vacuum and high gravitational forces (Ralston, 1993).

There are two distinct uses of the AUC: sedimentation equilibrium, SE, and sedimentation velocity, SV, both of which provide complimentary information about the size, shape and mass of a molecule. With SV a high angular velocity (usually above 20 k rpm) is used to sediment molecules which results in the depletion of solute from the meniscus and the formation of a sharp boundary between the depleted region and the body of the sedimenting solute which can be monitored by the AUC's optical system. The following two sub-sections give a general overview taken from a very good recent review by Laue and Stafford (1999) on the AUC. A practical guide for determining M by the AUC can be found in Byron (1996).

2.2.1 SEDIMENTATION VELOCITY

There are three forces acting on a sedimenting particle in solution: a sedimenting force F_s , which acts in the direction away from the centre of the rotor, and F_b and F_f (buoyant and frictional forces respectively) both of which oppose F_s .

$$F_s = m_h \omega^2 r_{bnd} \quad \text{Equation 2.9}$$

where ω is the rotational velocity (rad s^{-1}) of the centrifuge and r_{bnd} (cm), the distance between the molecule and the centre of the spinning rotor, which for practical purposes corresponds to the mid-point of the sedimenting boundary. The hydrated mass of the particle is

$$m_b = \frac{M_r}{N_A} (1 + \delta) \quad \text{Equation 2.10}$$

where δ is the hydration (or solvation) of the particle (in ml of solvent per gram of solute).

Opposing F_f is the frictional force F_f :

$$F_f = -\zeta v \quad \text{Equation 2.11}$$

where ζ is the frictional coefficient (g/s) and v the velocity (m/s) of the molecule; and buoyant force F_b :

$$F_b = -m_o \omega^2 r_{bnd} \quad \text{Equation 2.12}$$

The mass of fluid displaced by the sedimenting particle (m_o) is equivalent to $m \bar{v} \rho$ where \bar{v} is the partial specific volume¹³ of the sample (ml/g) and ρ the density of the solvent (g/ml). When the rotor is set spinning the sum of the three forces rapidly becomes equal to zero.

$$m_h \omega^2 r_{bnd} + (-\zeta v) + (-m_o \bar{v} \rho \omega^2 r_{bnd}) = 0$$

$$m_h \omega^2 r_{bnd} (1 - \bar{v} \rho) - \zeta v = 0$$

$$s^* = \frac{v}{\omega^2 r_{bnd}} = \frac{M_r (1 - \bar{v} \rho)}{N_A \zeta} \quad \text{Equation 2.13}$$

The sedimenting particle is described in terms of its apparent sedimentation coefficient, s , which is the velocity (v) of the molecule per unit gravitational acceleration ($\omega^2 r_{bnd}$) and has the unit of Svedberg (S), where 1 S is equivalent to 10^{-13} s. s can be calculated from monitoring the rate ($v = dr_{bnd}/dt$, where t is the time taken to travel a distance dr_{bnd}) of boundary movement, giving its apparent sedimentation coefficient, s^* :

$$s^* = \frac{d \ln r_{bnd}}{\omega^2 dt} \quad \text{Equation 2.14}$$

which is concentration dependent. By taking a series of measurements at different concentrations¹⁴, the value of s^* can be extrapolated to zero concentration, giving s^0 . This will reduce the non-ideality effects caused by other molecules. In order to use s^0 for comparative purposes, the measurements have to be standardised. Standard conditions refer to

¹³ the change in volume of a solution upon addition of 1 gram of solute.

¹⁴ Hydrodynamic non-ideality, which results from the counterflow of solvent from a sedimenting molecule thus slowing neighbouring particles is dependent on the concentration of the solution. Its effect is not thought to be significant for solutions consisting of spherical particles at concentrations of less than 5 mg/ml.

measurements taken at 20 °C in pure water ($20,w$) at infinite dilution. Measurements taken in a non-standard buffer (b) at a different temperature (T) can be corrected to standard conditions as follows

$$s_{20,w}^0 = s^0 \left(\frac{(1 - \bar{v} \rho)_{T,b}}{(1 - \bar{v} \rho)_{20,w}} \right) \left(\frac{\eta_{20,w}}{\eta_{T,b}} \right) \quad \text{Equation 2.15}$$

where η is the viscosity of the solvent (Poise). For a molecule of given shape, the more massive molecule will tend to sediment faster, while for a given mass, the more elongated molecule will sediment slower. The latter shape phenomenon is dependent on ζ which for shapes that have a larger surface area to volume ratio will have a higher ζ . Therefore for two molecules with the same volume, the one with a more rugged outer surface topology will have the higher ζ and a lower s .

2.2.2 SEDIMENTATION EQUILIBRIUM

Because in SV the rotor spins much faster than for SE studies, the effect of diffusion is minimal. However it can be noticed at the sedimenting boundary as a slope in the boundary i.e. if there was no diffusion, the boundary would be vertical. The rate of diffusion (or flux, j) as stated by Fick's first law depends on the concentration gradient i.e. so that

$$j = -D \frac{dC}{dx} \quad \text{Equation 2.16}$$

where D is the diffusion coefficient, C the solute concentration within a distance x ,

$$D = \frac{kT}{\zeta} \quad \text{Equation 2.17}$$

where k is the Boltzmann's constant and T the absolute temperature (K).

Besides stating that a high concentration gradient results in greater flux, the negative sign in equation 2.16 also indicates that molecules are transported away from the region of high solute concentration. This means that the high concentration of molecules present at the base of the cell as a result of sedimentation will cause those molecules to diffuse in a direction away from the base. This is more noticeable when F_s is not so strong i.e. when the rotor has been slowed down. The balance between sedimentation and diffusion results in a concentration

distribution ($d \ln C/dr^2$, where C is the concentration of the solute at a radial position r) along the length of the cell,

$$\frac{d \ln C}{dr^2} = \frac{\omega^2 M (1 - \bar{v} \rho)}{2kT} \quad \text{Equation 2.18}$$

which remains constant once thermodynamic equilibrium is established. The absolute molecular weight, M (gmol^{-1}), of a single molecular species for a homogenous solution or the weight average for a heterogeneous solution can thus be determined. The non-ideality of the solution, due predominantly to the presence of other molecules, has to be taken into account prior to reaching a final value for M . This is done by measuring the apparent M (M_{app}) at different concentrations. Then the actual M can be determined by extrapolating these values to infinite dilution.

The effect of non-ideality in SE measurements comes from (i) the volume occupied by the molecule, which is governed by its shape and flexibility, and (ii) the effect of charge-charge interaction between neighbouring molecules. If the interaction is repulsive, then both j and C increase, whereas if the forces are attractive, then the converse is true. Also at high concentrations, the effect of crowding can lower j as a result of non-specific interactions with other molecules. This can be partially suppressed with the addition of counter ions. The two contributions to non-ideality are additive and if not too severe can be expressed by the second virial coefficient B (ml mol/g^2). This term is derived from the second expansion term from a series expansion taking the concentration of solute to zero (see van Holde (1985)).

$$M_{\text{app}} \approx \frac{M_r}{(1 + 2BMC)} \quad \text{Equation 2.19}$$

where $2BM$ is the excluded volume i.e. the volume around a molecule which is unavailable to the centre of mass of an approaching molecule. Although solute solvent interactions do play a part in determining a value for B , (which is negative for repulsive interactions and positive for attractive interactions) its value is more likely to be governed by the size of the macromolecule i.e. its excluded volume. Larger molecules will generally have higher B values, as will rod shaped molecules (as opposed to sphere). The B values for random coils are very much dependent on the solvent.

The benefit of measuring M using AUC and SAS is that both are absolute methods of determination as opposed to the 'relative' methods which require calibration with macromolecules of known molecular weight such as size exclusion chromatography (SEC) and sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). However

mass can also be calculated from sequence information via mass spectrometry which like SDS-PAGE only requires small amounts of sample. While this a very precise method, it is also destructive and cannot be used to quantify self-association or association with other molecules. Therefore the key advantage of these solution techniques is to do with interpreting the gross size and shape of macromolecules and their assemblies in solution.

CHAPTER 3

THEORETICAL DETERMINATION OF SOLUTION PARAMETERS

This section reviews the development of solution modelling and how the solution parameters measured experimentally (see previous chapter) can also be calculated from simple geometric models. Figure 3.1 illustrates some of the different approaches that have been used to model particles in solution and will be referred to throughout this section. Previous reviews on this subject, namely from Garcia de la Torre and Bloomfield (1981), and Garcia de la Torre (1988) have been used as the primary source of reference for the following section.

3.1 WHOLE BODY APPROACH

One of the simplest ways of describing a molecule is as a sphere (Figure 3.1A), of volume, $V = M_r \bar{v} / N_A = 4/3 \pi \sigma^3$ (σ is the radius of the sphere and \bar{v} is the partial specific volume) and translation friction coefficient, $\zeta_0 = 6\pi\eta_0\sigma$, (where η_0 is the viscosity of the solvent (Poise)). For this model the radius of gyration, R_g (which for a sphere is 0.775σ), sedimentation coefficient, $s = M(1 - \bar{v}\rho) / N_A \zeta$ and diffusion coefficient, $D = kT / \zeta$ (where k is Boltzmann's constant and T is the absolute temperature) can be compared with those values measured experimentally. But the limited scope of using a single sphere particularly for extended molecules led researchers such as Simha (1940) and later, Scheraga and Mandelkern (1953) to derive formulae for the hydrodynamic properties of cylindrical rods and ellipsoids of revolution (Figure 3.1B).

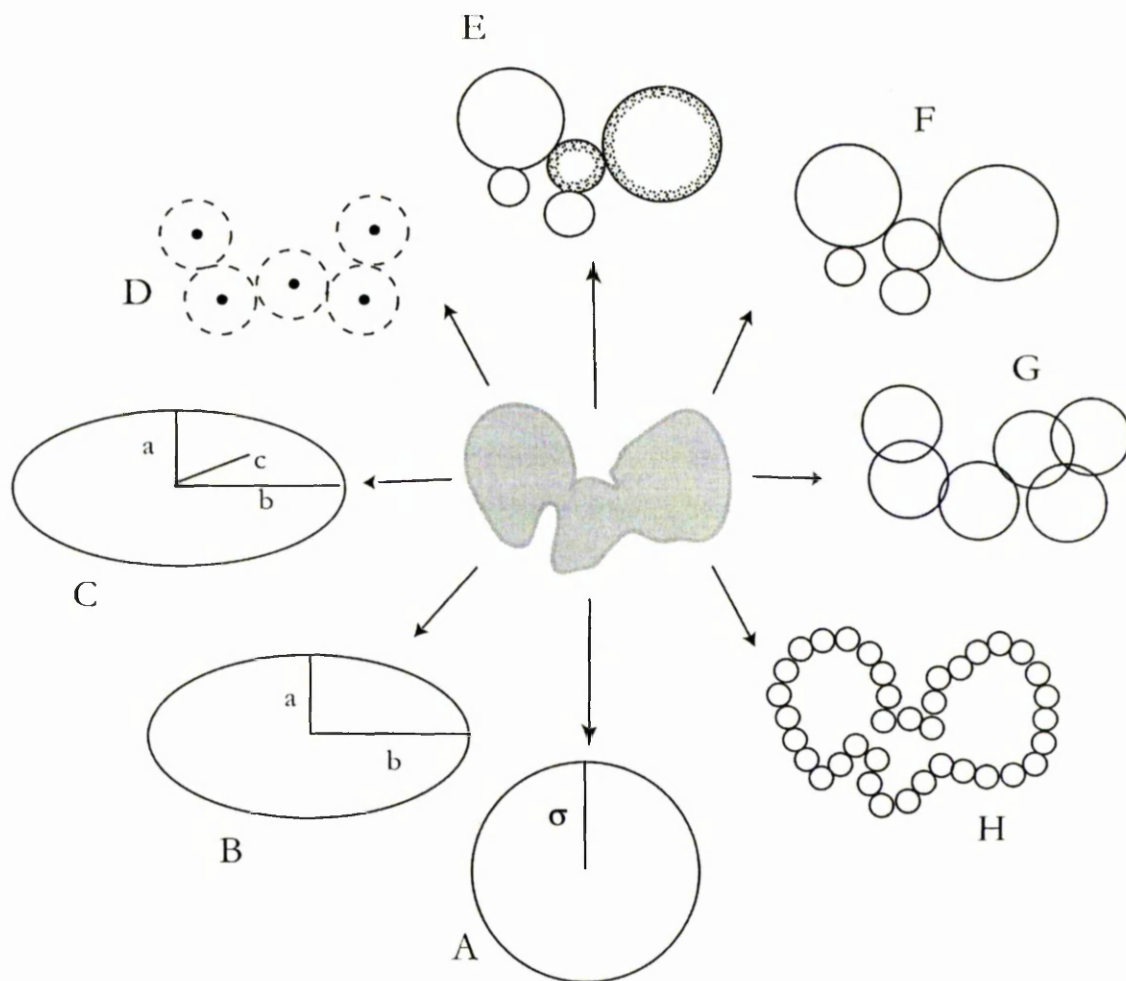


Figure 3.1 Eight representations of a molecule (depicted by the shaded object in the middle) from which hydrodynamic parameters can be calculated. Whole body representations such as the single sphere (A) with radius (σ), ellipsoid of revolution (B) and tri-axial ellipsoid (C) with two and three axial variables (a , b and c) respectively, are significantly simpler to describe than the multibody approaches D – H. The multibody (or bead) models correspond to different stages of model development. Beginning with each bead representing a frictional point on the molecule (D), to each bead representing a different sized frictional element. This initially involved covering the surface of each bead with many frictional points (E) (only two of the five beads illustrate the surface cover by friction points), but was later improved with analytical treatments for non-overlapping beads with variable size (F) and overlapping beads of equal size (G). As friction is only encountered at the molecular surface, very small beads placed around the molecule's surface are termed 'shell models' (H).

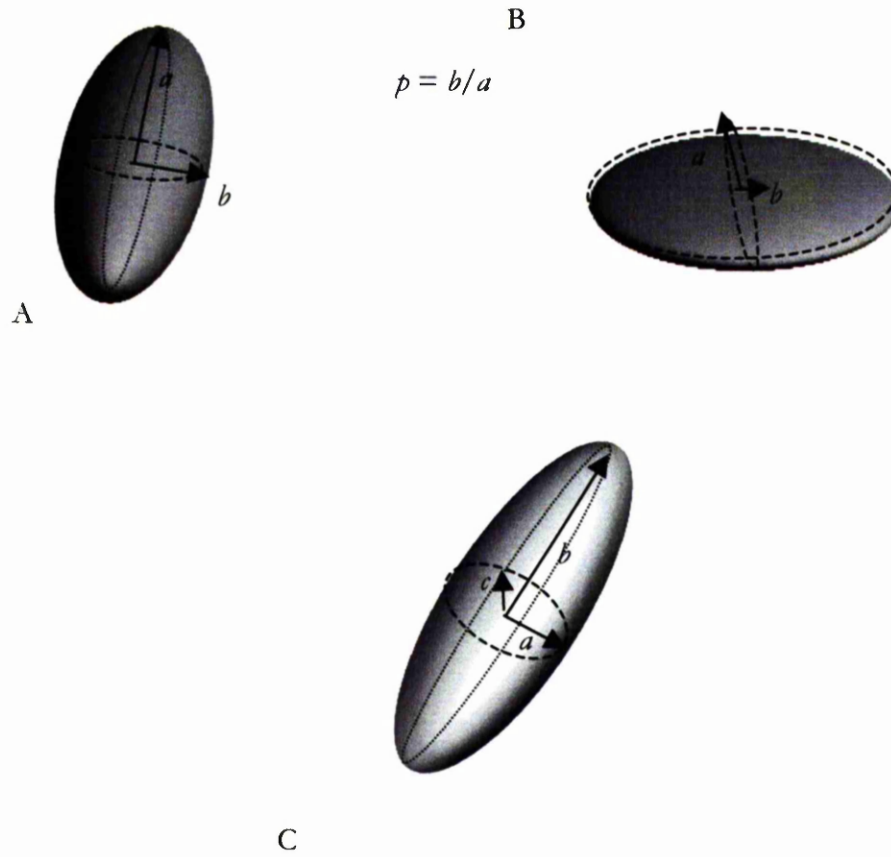


Figure 3.2 Three-dimensional models of (A) prolate and (B) oblate ellipsoids having two axis, a and b , where a denotes the semimajor axis and b the semiminor axis. While (C) the tri-axial ellipsoid has three axes – a , b and c . The expressions for both prolate and oblate ellipsoids are based on corrections reported by Koenig (1975) to Perrin's original paper (Perrin, 1934).

Ellipsoids for example, can be categorised as being either prolate ('cigar' shaped) or oblate ('disc' shaped) in shape. The degree of prolateness or oblateness is dependent on the $\zeta_e : \zeta_0$ ratio (also known as the Perrin factor) where ζ_e is the frictional coefficient of the ellipsoid. ζ_e can vary depending on the axial ratio p , (Teller *et al.*, 1979) where $p = b/a$, with b and a corresponding to the two axes running perpendicular to each other (Figure 3.2A and B). For a prolate ellipsoid, where $p < 1$,

$$\frac{\zeta_e}{\zeta_0} = \frac{(1 - p^2)^{1/2}}{p^{2/3} \ln \{ [1 + (1 - p^2)^{1/2}] / p \}} \quad \text{Equation 3.1}$$

whereas for an oblate ellipsoid, where $p > 1$,

$$\frac{\zeta_e}{\zeta_0} = \frac{(p^2 - 1)^{1/2}}{p^{2/3} \arctan(p^2 - 1)^{1/2}} \quad \text{Equation 3.2}$$

Differences between experimentally measured ζ and ζ_e could be attributed to a variety of different factors including poor shape description and the inability to determine the level of hydration. An improvement to the shape description was made by extending the ellipsoid model from two variable axis to three (tri-axial ellipsoid) (Figure 3.1C) (Harding and Rowe, 1983). The improvement in shape description meant that for those molecules where the shape is suitably described by the ellipsoid model, any deviation between predicted and experimental values could be attributed to differences in hydration.

$$\frac{\zeta}{\zeta_e} = \sqrt[3]{1 + \delta / \rho \bar{v}} \quad \text{Equation 3.3}$$

where δ is the hydration (grams of water per gram of protein)¹⁵. Computational strategies based on the whole body approach have the advantage of speed and simplicity in comparison to the multibody and finite element approach, both of which are also used to predict solution parameters.

3.2 MULTIBODY APPROACH

Although the tri-axial ellipsoid is an improvement on the single sphere and rod models, it does not adequately represent more complex structures such as the immunoglobulin G (IgG) class of molecules (Gregory *et al.*, 1987). One solution was to use bead models (molecules represented as an assembly of spheres), the theoretical foundation for which was the result of Kirkwood's (1954) generalised theory of diffusion. The theory postulates that irreversible processes acting on macromolecules in solution are determined by the hydrodynamic forces exerted on the solvent and the Brownian drift experienced by the molecule. Oseen (1927) had previously investigated forces acting on the solvent and from this derived the hydrodynamic interaction tensor, T to determine the effect each segment of a macromolecule had on the surrounding solvent flow pattern. Kirkwood used the bead model approach to model the behaviour of a single molecule within an unbounded system that contained a solvent with a molecular weight that was much lower than the actual molecule. For a molecule with N segments, each i^{th} segment is treated as being identical and is represented at the bead centre by a frictional element i.e. a point source of friction (Figure 3.1C). The frictional force (F)

¹⁵ The term hydration although used through out this thesis, is more accurately referred to as solvation as this term includes solvent species other than water that can also be trapped by or bound to the protein (Harding, 1989).

acting on the molecule is the sum of the forces acting on each of the segments (F_i) (Garcia and Bloomfield, 1981).

$$F_i = \sum_{i=1}^N \zeta_i (u_i - v_i) \quad \text{Equation 3.4}$$

where u_i is the velocity of the i^{th} element and v_i the velocity the solvent would have were that element not there. The movement of each segment creates an internal velocity field in the solvent so that the velocity of the solvent acting at the i^{th} element in the absence of other elements is v_i^0 , and is referred to as the unperturbed velocity. Using the interaction tensor (T) developed by Oseen (1927), which was later extended by Burgers (1938), Kirkwood was able to account for the perturbation caused by the other elements using the following equation.

$$v_i = v_i^0 - \sum_{\substack{j=1 \\ j \neq i}}^N T_{ij} F_j \quad \text{Equation 3.5}$$

where T is

$$T_{ij} = \frac{1}{8\pi\eta_0 R_{ij}} \left(I + \frac{r_{ij} r_{ij}}{R_{ij}^2} \right) \quad \text{Equation 3.6}$$

I is the unit tensor, and r_{ij} and R_{ij} are the distance vector (x, y, z coordinate system between i and j) and the distance between frictional elements i and j , respectively. Both Equation 3.4 and Equation 3.6 were amendments made by Garcia de la Torre (1981), from the original equations used by Kirkwood and Riseman (1949), Kirkwood (1954) and Bloomfield *et al.* (1967a) because of errors found in the original formalism. These errors resulted in the calculation of negative frictional coefficients (Zwanzig *et al.*, 1968) for some objects.

3.2.1 BEAD-FILLED MODELS

The limitation with the original Kirkwood formalism was that each element was treated as a regularly spaced point source of friction, which restricted the modelling of complex shaped molecules. By modelling each subunit as a surface shell of frictional points extrapolated to infinite numbers (Figure 3.1D), Bloomfield *et al.* (1967) were able to improve the hydrodynamic calculations based on the original Kirkwood derivation. This was shown by testing Bloomfield's approach with ellipsoids, the hydrodynamic parameters of which had already been solved by Perrin (Bloomfield *et al.*, 1967a). Even though the results were more

dependable than Kirkwood's (1954) approach, small differences were found between the Perrin equation for ellipsoids and that of the shell approach, the reason for which was not clear to the authors (Bloomfield *et al.*, 1967a).

Although the shell approach permits the modelling of subunits of different size, the problem with its implementation (as pointed out by Bloomfield *et al.* (1967) and Garcia de la Torre and Bloomfield (1977)) was that hydrodynamic forces were distributed on the subunit's surface as opposed to being localised at its centre. While this interpretation was more accurate in the sense that the effect of frictional forces due to solvent is encountered at the surface of the molecule, the actual formulation of the Oseen tensor had not taken this into account. This was particularly problematic when the size of the beads approached the overall size of the model. It took the modified interaction tensor developed by Rotne and Prager (1969), and Yamakawa (1970) (equation 3.7) to remedy the problem.

$$T_{ij} = \frac{1}{6\pi\eta_0\sigma} \left[\left(1 - \frac{9}{32} \frac{R_{ij}}{\sigma} \right) I + \frac{3}{32} \frac{r_{ij}r_{ij}}{\sigma R_{ij}} \right] \quad \text{Equation 3.7}$$

With this tensor, each element was given a finite size (σ), allowing those elements to overlap provided that they were of equal size (Figure 3.1G). This tensor was later generalised by Garcia de la Torre and Bloomfield (1977a) for elements of different radii, σ_i and σ_j (Equation 3.8) but was only valid for non-overlapping beads where $(\sigma_i + \sigma_j) \leq R_{ij}$ (Garcia and Bloomfield, 1981) (Figure 3.1F).

$$T_{ij} = \frac{1}{8\pi\eta_0 R_{ij}} \left[I + \frac{r_{ij}r_{ij}}{R_{ij}^2} + \frac{\sigma_i + \sigma_j}{R_{ij}^2} \left(\frac{I}{3} - \frac{r_{ij}r_{ij}}{R_{ij}^2} \right) \right] \quad \text{Equation 3.8}$$

The modified tensor used cubic substitution, where each of the two beads is substituted by a cubic array of spheres of radius, 0.4668σ , so that each bead contained 16 spheres (Garcia and Rodes, 1983). It was later noted by Garcia de la Torre and Rodes (1983) that this modified tensor was strictly only applicable to translational motion and not to rotational modes. They proposed the addition of a volume correction term which was found to yield better results particularly for rotational coefficients.

Recently, Zipper and Durchschlag (1997, 1998) have used equation 3.9 for overlapping beads of variable radii by temporarily treating the unequal spheres as having identical radii (σ):

$$\sigma = \sqrt[3]{\frac{\sigma_i^3 + \sigma_j^3}{2}} \quad \text{Equation 3.9}$$

Although there is no theoretical basis for this solution, the results obtained from this study were nearly identical to those models where overlaps had been removed. This solution was tested further by Carrasco *et al.* (1999) using only two beads (so that exact values for the hydrodynamic parameters could be derived theoretically) and a simplification of the original formula:

$$\sigma = \frac{\sigma_i + \sigma_j}{2} \quad \text{Equation 3.10}$$

They found very little difference in the final output when using either equation 3.9 or equation 3.10. The modified version was also found to perform better than when it had used equation 3.7, but the authors emphasise that overlapping bead models should be avoided when possible using the non-overlapping tensor given by equation Equation 3.8 instead. The latter advice is based on comparisons made by Carrasco and Garcia de la Torre (1999) on the different types of modelling approaches currently available. The approach most favoured by them is a modified version of Bloomfield's (Bloomfield *et al.*, 1967b) shell model (Figure 3.1G). The premise being that friction between solvent and molecule occurs only in those regions of the molecule exposed to the bulk solvent i.e. at the surface. So only those beads at the surface will contribute to the frictional properties of the model.

3.2.2 SHELL MODELS

Bloomfield *et al.* (1967) used very small beads to outline the surface rugosity of certain multisubunit proteins and viruses, and also simpler shapes such as spheres (Figure 3.1E), ellipsoids and cylinders. The hydrodynamic parameters determined by this approach was reasonably accurate for simple shaped objects, but less so for biological complexes. The discrepancy for objects such as rotational ellipsoids was found to be caused by the error in the original Kirkwood formula (Swanson *et al.*, 1978). While this will have contributed to the results determined for the biological complex, it is also likely that the starting model, which was based on an electron microscopy image, was not representative of the actual solution structure of the protein or virus (Bloomfield *et al.*, 1967b). The shell approach used by Teller *et al.* (1979) was further improved by using atomic data to position non-overlapping beads of 1.4 Å radius (comparable to the size of water molecules) on non-hydrogen atoms around the protein surface. The frictional coefficients were calculated using both the original Kirkwood-Oseen tensor and the modified tensor. The former method was found to yield frictional coefficients greater than the experimentally observed values while the latter was found to be less. However the best correlation was found when spheres were only placed on the charged groups of surface atoms with only a single hydration layer. The latest advance in shell

modelling is embodied in the computer program *HYDROPRO* (Garcia de la Torre *et al.*, 2000), which generates shell models of proteins from PDB (protein data bank) formatted files (Berman *et al.*, 2000). This approach gave reasonably accurate predictions for the hydrodynamic parameters of thirteen pre-characterised test proteins (Garcia de la Torre *et al.*, 2000).

To summarise the main points of this section, there are two types of interaction tensors for determining the hydrodynamic properties of a bead model. The constraints on model building placed by both tensors are that all beads must have equal radii if beads overlap or alternatively no overlaps for models with differently sized beads. For the latter case the distance between two beads must be equal to or greater than their radii. The *ad hoc* procedure proposed by Zipper and Durchschlag (1997) extended this procedure for overlapping beads with variable radii. Provided that these constraints are followed the only other limitations to the model building procedure is accuracy (in terms of minimising the difference between experimental and calculated values for a protein) and efficiency.

Efficiency can be gauged by the speed with which the hydrodynamic parameters are calculated and the provision of features that facilitate the modelling procedure. While the latter is of particular concern due to the lack of appropriate editing features in most popular molecular visualisation tools (see Table 4.3), the former can be reduced by using better computer hardware; fewer beads in the model and using approximate methods for hydrodynamic calculations. Garcia de la Torre (1988) has reviewed two basic approaches for hydrodynamic calculations of rigid particles using bead models: the rigorous (more accurate) treatment and the approximate (much faster) double sum method. The difference in speed is due primarily to the procedure used to calculate the “supermatrix” (this describes the motion of each bead), which is roughly proportional to the number of beads (N) in the model. For the double sum method, this is approximately N^2 , whereas for the rigorous approach it is N^3 . To put this into perspective, the difference in speed for a model composed of 500 beads run on a Silicon Graphics MIPS R1000 180 MHz CPU (central processing unit) is approximately 1000 fold, taking only a couple of seconds when using the approximate method (Carrasco and Garcia de la Torre, 1999a). Programs used for hydrodynamic calculations of rigid bodies include *HYDRO* (Garcia *et al.*, 1997), *SOLPRO* (Garcia *et al.*, 1994) and *COEFF* which offers three different routines: *SUPC*, *GS*, *DLAG* (Spotorno *et al.*, 1997). For segmentally flexible bodies with two subunits, the programs *FLEX* and *SUPFLEX* are also available (Spotorno *et al.*, 1997).

3.3 HYDRODYNAMIC TREATMENT FOR RIGID BODIES

The supermatrix, \mathcal{Q}_{ij} is an $N \times N$ matrix ($i = j$, $T_{ii} = 1/\zeta_i$ $i \neq j$, $T_{ii} = \mathcal{Q}_{ij}$) corresponding to rotation, translation, shear flow. It is related to the frictional force (F) in the following way:

$$\sum_{j=1}^N \mathcal{Q}_{ij} F_j = \zeta_i (u_i - v_i^0) \quad \text{Equation 3.11}$$

where \mathcal{Q}_{ij} is

$$\mathcal{Q}_{ij} = \delta_{ij} I + (1 - \delta_{ij}) \zeta_j T_{ij} \quad \text{Equation 3.12}$$

where δ_{ln} is the Kroenecker delta ($l = n$, $\delta_{ln} = 1$; $l \neq n$, $\delta_{ln} = 0$). Equation 3.12 is related to equation 3.4, which if taking the perturbation term (Equation 3.5) into account gives:

$$F_i = \zeta_i (u_i - v_i^0) - \zeta_i \sum_{\substack{j=1 \\ i \neq j}}^N T_{ij} F_j \quad \text{Equation 3.13}$$

Re-writing equation 3.13 in a more compact form gives

$$B\mathfrak{S} = \mathfrak{u} \quad \text{Equation 3.14}$$

where \mathfrak{u} and \mathfrak{S} describe a six-dimensional velocity and force vector respectively for the six degrees of freedom available to a rigid particle i.e. three for translation and three for rotation. B is a $3N \times 3N$ supermatrix containing the $N \times N$ matrix, \mathcal{Q}_{ij} . The inverse of B is the 6×6 friction tensor Ξ , which is written as:

$$\Xi = B^{-1} \quad \text{Equation 3.15}$$

The relationship between Ξ and \mathfrak{S} is linear provided that the particle is moving in a fluid with low Reynolds number¹⁶ (Garcia de la Torre, 1989). Inserting equation 3.15 into equation 3.14 gives the standard expression for the frictional force (\mathfrak{S}) acting on translating particles:

$$\mathfrak{S} = \Xi \mathfrak{u} \quad \text{Equation 3.16}$$

and for a rotating particle the torque (Γ) exerted from a reference point R on the particle is:

¹⁶ The Reynolds number (Re) is a characteristic of a fluid flow pattern. Fluids with identical Re will have the same pattern. $Re = \rho v l / \eta$ where ρ is the density of the fluid with viscosity η , moving with velocity v relative to a solid characterised by linear dimension l .

$$\Gamma_R = \Xi_{R,r} \omega \quad \text{Equation 3.17}$$

Where ω is the angular velocity and, Ξ_r and $\Xi_{R,r}$ are the translational friction and rotation tensor respectively. $\Xi_{R,r}$ unlike Ξ_r can change depending on the position of the reference point (R). The volume correction introduced by Garcia de la Torre and Rodes (1983) is used to improve the accuracy for calculating $\Xi_{R,r}$.

$$\Xi_{R,r} = \Xi_{R,r} + 6\eta_0 V/I \quad \text{Equation 3.18}$$

The generalised Einstein relationship between diffusion (expressed as the diffusion matrix \wp) and Ξ is:

$$\wp = kT\Xi^{-1} \quad \text{Equation 3.19}$$

where \wp is either rotational diffusion (\wp_r) if $\Xi_{R,r}$ is used or translational diffusion (\wp_t) if Ξ_r is used. T and k are the absolute temperature and Boltzmann constant, respectively. This equation can be expanded to give a 6 x 6 matrix that can be divided into four 3 x 3 blocks as follows:

$$\begin{bmatrix} \wp_{tt} & \wp_{tr}^\Theta \\ \wp_{tr} & \wp_{rr} \end{bmatrix} = kT \begin{bmatrix} \Xi_{tt} & \Xi_{tr}^\Theta \\ \Xi_{tr} & \Xi_{rr} \end{bmatrix}^{-1} \quad \text{Equation 3.20}$$

where Θ denotes transposition and tt , rr and tr are translational, rotational and the coupling of translation and rotation respectively. The translational diffusion coefficient (D_t) can be determined directly from the trace (Tr) of the \wp_{tt}

$$D_t = \frac{1}{3} Tr \wp_{tt} \quad \text{Equation 3.21}$$

and similarly for rotational diffusion coefficient (D_r), from the \wp_{rr} block (Garcia de la Torre and Bloomfield, 1981). Some of these parameters are particularly sensitive to experimental conditions. For instance, ζ_t is only sensitive to η_0 , whereas D_t and D_r are both sensitive to η_0 and temperature. A joint parameter, present in the derivation for s , D and ζ is the Stokes radius, R_s . This is essentially the rearrangement of σ in the friction equation and is used later in chapter 5 as an alternative indicator for s , D_t and ζ_t .

Other parameters that can also be derived are the rotational relaxation times, τ obtained from the eigen values of the ϕ_{rr} ; R_g , which is:

$$R_g = \frac{\sum_i m_i r_{i0}^2}{\sum_i m_i} \quad \text{Equation 3.22}$$

where r_{i0} is the distance between the coordinate i of each bead and the centre of mass for the model and the intrinsic viscosity, $[\eta]$.

$[\eta]$ like D_f and τ is sensitive to the shape (and hence hydration) of the molecule. Like solvent viscosity, η_0 it can be described in terms of its contribution to resisting solvent flow. This actually describes the specific viscosity of the molecule, η_{sp} but extrapolating this to infinite dilution will give $[\eta]$. It can be measured experimentally by the time taken for a volume of macromolecular solution to travel a certain distance (as in the Ostwald viscometer) or from the energy released as result of shearing with other molecules and with the solvent. Harding (1998) has published a comprehensive review of $[\eta]$, with a recent update for its calculation from bead models given by García de la Torre and Carrasco (1998) from which the details presented below were taken. A volume correction to $[\eta]$ which was not present in the initial derivation by Riseman and Kirkwood (1950) is required to predict its value more accurately. The problem arises from the García de la Torre and Bloomfield (1978) derivation for $[\eta]$ which when applied to a single bead, its r_i^{17} is zero which means that $[\eta]$ is also zero. While this is unlikely to be a problem with bead models comprised of many beads, the remedy was to substitute each bead with a cubic array of smaller spheres, a strategy that was similar to the shell approach used by Bloomfield *et al.* (1967). Although this resolved the problem of the single bead (albeit by increasing the computation time), it required an additional volume correction factor that was equivalent to $[\eta]$ for a single bead monomer i.e.

$$\eta_1 = \frac{5N_A V_1}{2M_{r1}} \quad \text{Equation 3.23}$$

where V_1 is the total volume of the entire model

$$V_1 = \frac{4}{3}\pi \sum_{i=1}^N \sigma_i^3 \quad \text{Equation 3.24}$$

and M_{r1} the molecular mass of the single bead, so that $\eta = \eta_{corr} + \eta_1$.

¹⁷ in this case r_i is a position vector for bead i with respect to a particular viscosity centre which for centrosymmetric particles coincides with its symmetry

Because shell models only cover the surface of the model, the sum of the bead volumes is unlikely to correspond with those of bead filled models, where relatively fewer beads are used. Therefore the recommendation is that the volume correction is to be used when only a few beads are present in the model. Zipper and Durchschlag (2000) using as few as 44 beads also avoided the correction factor in their model calculations. However they also used the *ad hoc* modification of the interaction tensor for calculating parameters from models containing overlapping beads of different size (Zipper and Durchschlag, 1998).

3.4 SCATTERING PROPERTIES FROM BEAD MODELS

The constraints placed on bead models (also known as Debye sphere models when used in the context of SAS) for making hydrodynamic predictions are different for SAS. The emphasis of the former method is that each bead acts as a source of friction, whereas with scattering, each bead acts as a point scatterer. While friction is generally encountered at the surface of the molecule, scattering is dependent on both the shape and internal structure of the molecule. In this respect, it is important that the mass distribution of the molecule is modelled accurately. So, for example a shell model is considered unsuitable, unless it is filled with beads. This is precisely how *HYDROPRO* (Garcia de la Torre *et al.*, 2000) works i.e. by filling the interior of the shell model with equivalent sized non overlapping beads. Other methods include the grid based approach which has several permutations, and the running mean method.

The grid based approach, where beads are placed either at the centre of the cell (Mayans *et al.*, 1995; Perkins, 1989) (for those cubes possessing more than a set number of atomic coordinates); at the α -carbon position of the residue (Perkins *et al.*, 1998); at the centre of gravity (c.o.g) of those residues (or atoms) within each cube (Byron, 1997) (Zipper and Durchschlag, 2000). Whereas in the running mean method used by Zipper and Durchschlag (2000), a set number of neighbouring atoms along the protein chain or sufficient atom numbers to fill a bead of set size are grouped to form a single bead. The volumes of each bead are adjusted so that their total volume represents the volume of the anhydrous molecule. For SAXS and hydrodynamic calculations, hydration is added to the overall volume, by (for example) enlarging the beads to take into account a single hydration layer around the surface of the molecule. A recent comparative study by Zipper and Durchschlag (2000) on the grid based and running mean method showed both methods to give reasonably accurate hydrodynamic and SAS predictions.

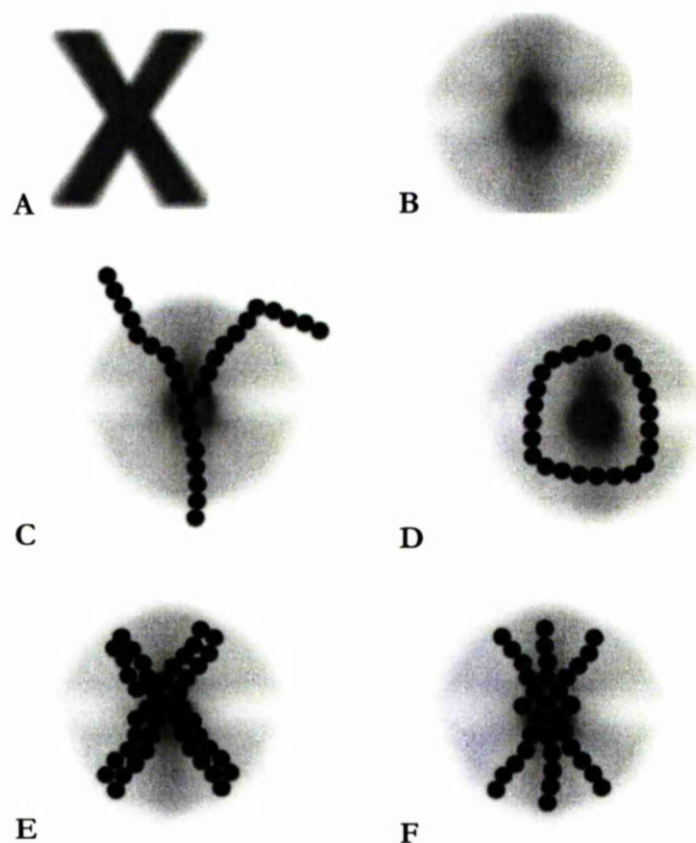


Figure 3.3 An image of a molecule (A) rotated by a variety of different orientations to give an impression of how the scattering elements are distributed within that molecule. C and D are two unlikely examples of bead models that would not correspond to the scattering curve generated by A, while E and F would. This illustrates how, scattering is unlikely to provide a unique solution for a model but can infer the unlikelihood of certain models.

Because each bead represents a scattering centre, similar to the model presented in Figure 3.1D (where each bead instead acts as a frictional centre), the scattering caused by incident radiation 'hitting' these scatterers is used to determine their general distribution inside the molecule or PDDF¹⁸ (probability distance distribution function, $p(r)$). The application of this function is illustrated in Figure 3.3, where a molecule represented by a cross is placed alongside an impression of that image under various different orientations. There are clearly a variety of bead models that would (Figure 3.3 E and F) and would not (Figure 3.3 C and D) fit this image, giving this function the potential to discriminate between unsuitable models.

¹⁸ PDDF can be described in terms of the number of lines of distance r connecting scattering elements within the model.

$p(r)$ can be determined from the angular variation of the intensity of the scattered radiation using equation 3.25 which was based on equation 2.1 from the previous chapter.

$$F(Q) = \int_0^{\infty} p(r) \frac{\sin(rQ)}{rQ} dr \quad \text{Equation 3.25}$$

where $p(r)$ is expressed as a Fourier transform (Perkins, 1988b):

$$p(r) = \frac{1}{2\pi^2} \int_0^{\infty} (rQ) I(Q) \sin(rQ) dQ \quad \text{Equation 3.26}$$

which can extend from $0 \leq Q \leq \infty$. However, in practice, the integrals over distance can be extended to an upper limit $r = D_{max}$ where D_{max} is the longest distance between two point scatterers (if $r > D_{max}$ or if $r = 0$, then $p(r) = 0$). $p(r)$ generally increases with r , until r exceeds $0.5D_{max}$ of the model where $p(r)$ may then fluctuate depending on the shape of the molecule (Glatter, 1991). $p(r)$ is calculated by the distribution of distances, r between two point scatterers, N^2 . For a 3-D model this will mean the distances will be cubed and so N^2 will be proportional to r^6 (Garcia *et al.*, 1999).

For bead models (where each bead has a finite size), $p(r)$ can be determined from the distribution of distances between points of a sphere with radius σ (P_{11})

$$P_{11}(r) = \frac{3}{16\sigma^6} r^2 (r - 2\sigma)^2 (r + 4\sigma) \quad \text{Equation 3.27}$$

and the distribution of distances between two beads (1 and 2), separated by R_{12} having the respective radii of σ_1 and σ_2 .

$$P_{12}(r) = \frac{3r}{160R_{12}\sigma_1^3\sigma_2^3} (\sigma_1 + \sigma_2 + a)^3 \quad \text{Equation 3.28}$$

$$(-4\sigma_1^2 + 12\sigma_1\sigma_2 - 4\sigma_2^2 - 3\sigma_1a - 3\sigma_2a + a^2)$$

so if $R_{12} - \sigma_1 - \sigma_2 < r < R_{12} + \sigma_1 - \sigma_2$

$$P_{12}(r) = \frac{3r}{20R_{12}\sigma_1^3\sigma_2^3} (\sigma_1)^3 (-\sigma_1^2 + 5\sigma_2^2 - 5a^2) \quad \text{Equation 3.29}$$

or if $R_{12} + \sigma_1 - \sigma_2 < r < R_{12} - \sigma_1 - \sigma_2$, or

$$P_{12}(r) = \frac{3r}{160R_{12}\sigma_1^3\sigma_2^3}(\sigma_1 + \sigma_2 - a)^3$$

$$(-4\sigma_1^2 + 12\sigma_1\sigma_2 - 4\sigma_2^2 - 3\sigma_1a - 3\sigma_2a + a^2) \quad \text{Equation 3.30}$$

or if $R_{12} - \sigma_1 + \sigma_2 < r < R_{12} + \sigma_1 + \sigma_2$.

$$p(r) = \sum_{i=1}^N fV_i^2 P_{11}(r; \sigma_i) + 2 \sum_{i \neq j}^N \sum_{j=1}^N fV_i fV_j P_{12}(r; \sigma_i; \sigma_j; R_{12}) \quad \text{Equation 3.31}$$

where $a = r - R_{12}$ and $p(r)$, $fV_i = V_i / \sum V_i = \sigma_i / \sum \sigma_i^3$ is the volume fraction of the bead. It assumes that all beads are non-overlapping and possess a uniform scattering density.

$p(r)$ can also be used to calculate R_g and $I(0)$ using the whole of the scattering curve as opposed to just the Guinier region. The values obtained by this approach may differ slightly with the Guinier approach due to the inclusion of the high Q part of the curve which will have a reduced signal to noise ratio.

$$R_g^2 = \frac{1}{2} \int_0^L p(r) r^2 dr \quad \text{Equation 3.32}$$

$$I(0) = 4\pi \int_0^L p(r) dr \quad \text{Equation 3.33}$$

R_g can also be calculated from the crystallographic coordinates using:

$$R_g^2 = \frac{\sum_{i=1}^n f_i R_{i0}^2}{\sum_{i=1}^n f_i} \quad \text{Equation 3.34}$$

where f is the scattering length from each atom (in this case for X-rays) and R_{i0} the distance between the coordinate and the centre of the scattering particle. The R_g can also be determined for overlapping beads from:

$$R_g^2 = \sum_{i=1}^n \left(\frac{R_{gi}^2 + R_{i0}^2}{n} \right) \quad \text{Equation 3.35}$$

where R_{gi} is the R_g for each of the n units.

Both $p(r)$ and $I(Q) \text{ v } Q$ curves calculated from bead models can be compared with experimental SAS measurements. Their importance in this chapter is predominantly due to their relevance in describing the shape and conformation of macromolecules. Garcia de la Torre *et al.* (1999) recently proposed combining both R_g and D_{max} to create a new shape function $H(D_{max}/R_g)$. Such functions are more commonly associated with ellipsoidal models which, although not mentioned in this section, can also be used to calculate ζ and hence s and D (Kumosinski and Pessen, 1982) and also to simulate the structure and thermodynamic properties of protein solutions (Sjoberg and Mortensen, 1997). The calculations described in this chapter are those implemented in the *SOLPRO* program (Garcia *et al.*, 1999). However this program is only one of several other programs that includes *GNOM* (Svergun, 1992), *SCT* (Mayans *et al.*, 1995) and *CRY SOL* (Svergun *et al.*, 1995). The latter two programs predict $\ln I(Q) \text{ versus } Q$ scattering curves, whereas *GNOM* predicts the PDDF from the $I(Q) \text{ versus } Q$ scattering curve.

CHAPTER 4

ANALYSIS AND IMPLEMENTATION OF SOFTWARE REQUIREMENTS

The various aspects of analysis and implementation of software are dealt by the broad heading of software development or software engineering. Like other engineering topics, software engineering is concerned with providing effective solutions to tasks requiring knowledge specific to computers and computing. Engineers have formalised this approach by (i) understanding the task and the problems associated to it; (ii) defining (or re-defining) those requirements that will offer a solution and then (iii) implementing them in a computer program. The program can then be (iv) tested to ensure that the implementations have satisfied the original requirements. These four steps usually follow a refinement procedure that involves repeating step (ii) or (iii) until the testing is satisfactory.

Chapter 1 briefly introduced some of the more popular strategies currently available for structural studies of macromolecules and the benefits of solution modelling techniques for studying complex interacting systems. Chapter 2 and 3 have covered the necessary background information required for understanding this particular subject area as identified by step (i). This chapter describes the features that have been implemented in *SOMO* (SOlution MOdeller), a program for generating bead models from atomic coordinate data. *SOMO* is a single module belonging to *MOLed* (MOlecular editor), a molecular modelling software package that is being developed by Dr Ron Poet at the University of Glasgow to model molecular interactions using biophysical and atomic data and to provide a platform for further software development.

The apparent ease of software development can result in the duplication of features in some computer programs. This is the main reason for the inclusion of libraries in programs. Libraries are usually a collection of small program routines or modules that offer solutions to a set of problems that can be encountered in vastly different projects. A good library will not only be easy to use but will also be flexible for use by different programs. Object oriented languages such as C++ facilitate the coding of computer programs as small modules for incorporation into libraries. *MOLed* is a collection of library modules written for molecular modelling. Its design is briefly covered in section 4.1.

Section 4.2 gives an outline of the *Trans* (Spotorno *et al.*, 1997) and *AtoB* (Byron, 1997) algorithms; describing some of the features that have been used to optimise each algorithm. This is an important aspect of program design as it allows permutations of a particular algorithm to be explored and also means that different bead model generation methods can be

used depending upon the requirements of the study. For instance, *SOMO* associates a certain number of water molecules with each residue (Kuntz and Kauzmann, 1974). But this number can be changed by the user in response to different experimental conditions that can effect the number of bound water (such as temperature and solvent composition).

Section 4.3 deals with the graphical interface and covers the different approaches that were taken to provide an effective modelling interface. The bead coordinates generated by *SOMO* can be displayed either directly from the *MOLed* interface or by writing the coordinates to a file that can be read by other molecular visualisation packages such as *Rasmol* (Sayle and Milner-White, 1995) and *VMD* (Humphrey *et al.*, 1996).

4.1 PROGRAM DESIGN

A typical cycle for software development especially in the scientific community involves finding a problem that can be addressed computationally and then writing a program that will solve it. These programs are usually small and perform relatively simple specific tasks both correctly and efficiently. It is a simple, workable development cycle, which can run into problems if substantial modifications are made to the program. In such an event the possibility of re-writing the program must be considered. The advantage of re-writing the program is that more emphasis can be given to its design so that further modifications are easier to implement. However if the program has already been extended quite considerably and demand for its use is not so great, then the prospect of re-writing seems less justifiable due to the time and effort that would be required for such an undertaking.

The *MOLed* package comprises a collection of program libraries used previously by 1D-3D (Milnerwhite *et al.*, 1990), a molecular visualisation program specifically designed for studying intra-molecular hydrogen bonding (Belhadjmostefa *et al.*, 1991). All these libraries are written in C++, a programming language that was developed from the C programming language (Stroustrup, 1995) with added features for object oriented (OO) development. Unlike traditional procedural languages such as C and FORTRAN, where a program is essentially a collection of functions; with OO programs some of those functions are grouped together as objects that interact with other objects to form a

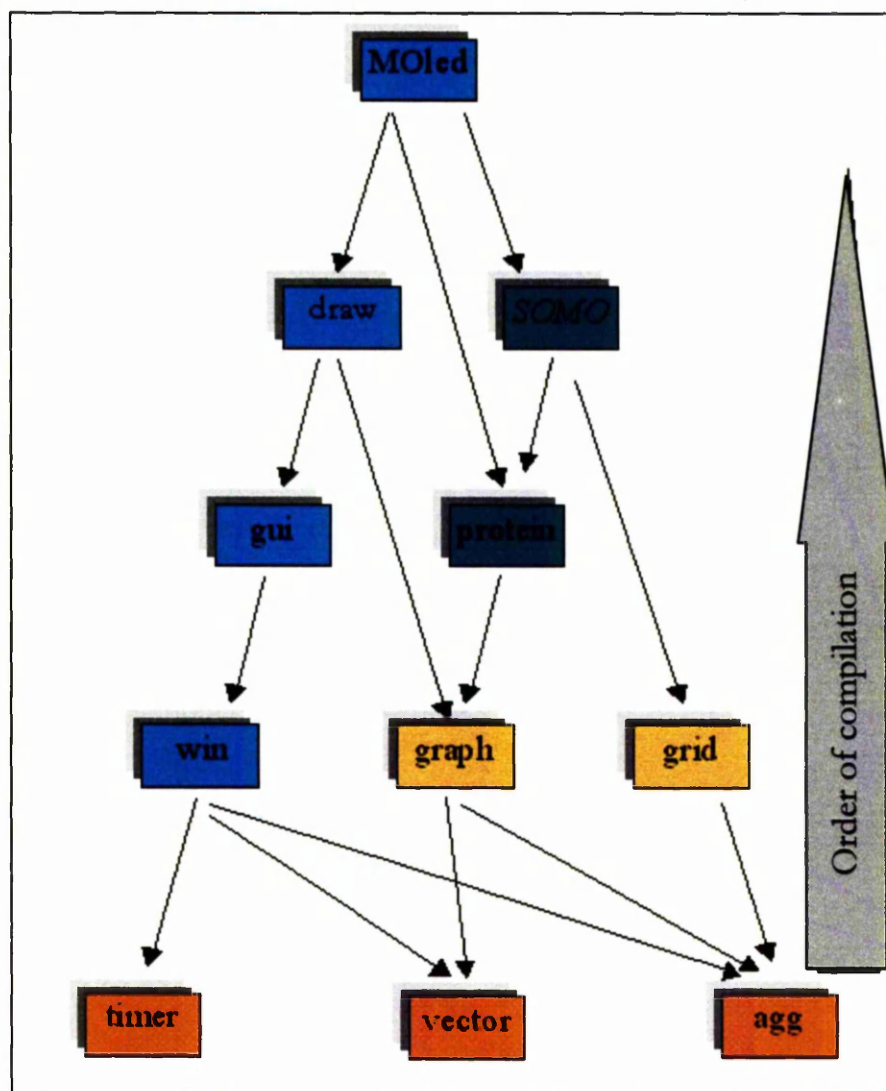


Figure 4.1 The arrangement and dependencies of the four different types of libraries that constitute MOled are: **Utility**, **Article**, **Base** and **Display**. The dependencies shown by the connecting arrows provide the order in which the libraries have to be compiled. For example, the *protein* library is dependent on the *graph* library, which in turn is dependent on both the *vector* and *agg* libraries. So *protein* can only be compiled once *vector*, *agg* and *graph* have been successfully compiled in that order.

functional program. OO programs share the same design philosophy as those used to develop program libraries i.e. building a large block of program code from lots of smaller self-contained blocks of code. C++ is a flexible, powerful language that is available for post modern computer systems and is used for developing a significant proportion of modern molecular modelling software.

MOLed currently consists of eleven libraries written in OO-C++, with SOMO being one of them. Each library consists of a collection of functions that are categorised into four groups for performing specific tasks as illustrated in Figure 4.1.

These groups are the:

- Base libraries, which contain general functions such as vector calculations and timer events that are likely to be required by the other libraries.
- Article libraries, possessing functions for reading, processing and storing information obtained from PDB files (Berman *et al.*, 2000).
- Utility libraries, providing more generic processing facilities compared to those contained in the article libraries and so are of use to more than one class of library.
- Display libraries, which are responsible for the graphical representation and display of the molecule.

SOMO belongs to the article group of libraries because it reads in the atomic details of a molecule in PDB format and stores this information for processing into bead models. Information about each bead such as its position, radius, mass, surface accessibility and amino-acid residue details was stored in the bead object of *SOMO*. Likewise, atomic data was stored in the RPprotein object in the protein module. Each bead (and hence the bead object) was arranged in a sequential order that was stored in the BeadsSeq object. This object also holds the RPprotein object. The BeadsSeq object provides an easy means of accessing molecular information about each bead and also correlates the atomic coordinate model with its bead model counterpart. The latter is best illustrated by Figure 4.2, in which an early version of *MOLed* is used to display three structural representations of a molecule.

The hierarchical arrangement of the *MOLed* libraries governs the order in which they should be compiled. For instance graph can only be compiled once vector and agg have been compiled. So *SOMO* would require vector, agg, graph, grid and protein to be compiled in that order. The features of each of these libraries are described briefly in Appendix 1. Because of the benefits of the C++ language and the flexibility of the *MOLed* libraries, it seemed appropriate to incorporate these libraries into a single OO-C++ program – *SOMO*, that would benefit from the features and algorithms available in both *AtoB* and *Trans*.

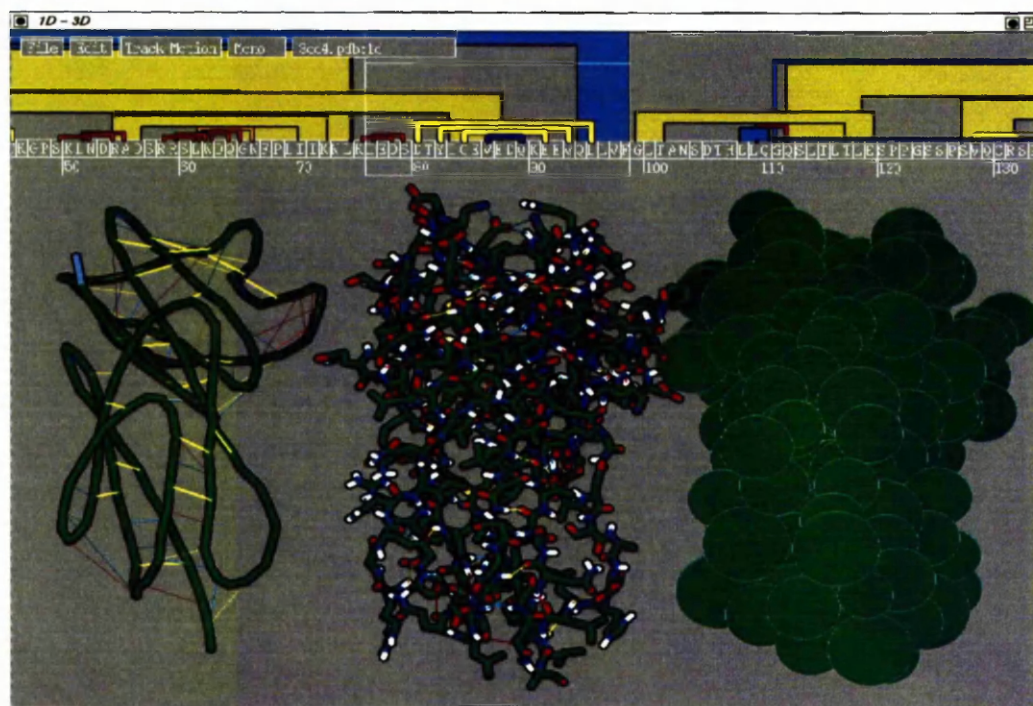


Figure 4.2 Bead model displayed alongside sequence, chain and atom representation, on the 1D-3D interface.

4.2 ALGORITHM OPTIMISATION

There are at least six research groups (listed in Table 4.1) that have in the last five years developed computer programs for generating bead models. These programs can be divided into two general categories. The four highlighted programs generate hydrodynamic bead models from high resolution atomic data. Those programs that are not highlighted generate (or restore) a *de novo* bead model from either SANS or SAXS scattering curve. The bead models generated by all these programs have been used to characterise the gross structural conformation of a variety of different biological systems (see program references for examples). Therefore the prospect of improving bead modelling algorithms is of considerable interest for those involved in such studies.

SOMO focuses on the development of the highlighted group of programs. It currently offers two distinct methods for generating bead models, using either the *AtoB* or *Trans* algorithm as its basis. The primary reason for choosing these particular algorithms was because they have both been successfully applied in solution studies for a variety of different systems and embody two very different approaches for generating bead models (Byron, 2000). The *Trans*

program also possesses features for processing bead models that are not only beneficial to the *AtoB* algorithm but potentially for other types of bead generating algorithms.

Author(s)	Software
Chacon <i>et al.</i> (2000)	<i>Dalai-GA</i>
Svergun <i>et al.</i> (1999)	<i>DAMMIN</i>
Garcia de la Torre <i>et al.</i> (2000)	<i>HYDROPRO</i>
Spotorno <i>et al.</i> (1997)	<i>Trans</i>
Byron (1997)	<i>AtoB</i>

Table 4.1 List of recent programs for generating bead models from PDB files (shaded) and neutron and X-ray scattering data. Not included in this list, because they have no program listing, are permutations on the methods just described by Banachowicz *et al.* (2000); Perkins *et al.* (1998) and Zipper and Durchschlag (1998).

Both *AtoB* and *Trans* generate models whose calculated parameters are close to experimental values, which is why they have been applied successfully with different experimental systems. However there is still scope to improve these algorithms by bringing experimental and predicted parameters even closer together. Traditionally improvements to the algorithm were made by altering the program code and testing the modified algorithm with molecules that have been extensively characterised by crystallographic and solution techniques. *Trans* also offers a set of basic processing operators that allows the user to alter the model by making limited changes to the algorithm without having to alter the programs infrastructure. The algorithm is evaluated by comparing the closeness of the newly generated biophysical parameters with previously determined experimental parameters. A close correlation between experimental and theoretical parameters suggests an improved algorithm. The same principles also apply to *SOMO*, but extra operators and an input data table that allows bead packing and hydration to be altered without re-compilation, offers even more flexibility for altering the basic algorithm.

The procedure for generating and altering bead models from atomic coordinates using *SOMO* involves running a series of different programs. The first step is to obtain atomic data in the PDB file format. This is obtained either experimentally by X-ray crystallography or NMR, or in the absence of experimental data by predictive methods, which predominantly use primary sequence data. Having generated a PDB file of the structure(s) under study, the next step is to discriminate those residues on the surface of the molecule which are exposed to the solvent

environment from those that are buried. This is done using *ASA* (Accessible Surface Area) (Spotorno *et al.*, 1997), a part of the *BEAMS* suite of programs which uses the rolling sphere (Lee and Richards, 1971) algorithm to evaluate the surface exposition of each atom. Based on an user-selected cut-off value, *ASA* assigns an arbitrary value of 1 to the “colour” field to indicate that the atom belongs to the side chain of an exposed surface residue, 6 if the exposed atom is not part of a side chain, and 10 for any atom that is not surface accessible, i.e. buried. *SOMO* sorts residues according to their surface accessibility to ensure that the surface topology of the bead model is consistent with the atomic surface representation.

ASA reformats the PDB file for *SOMO* which then generates bead co-ordinate files that can be read by molecular visualisation programs to display the models or by programs such as *HYDRO* (Garcia *et al.*, 1994) and *COEFF* (Spotorno *et al.*, 1997) to calculate theoretical hydrodynamic parameters. The bead models generated by *SOMO* are based on the original *AtoB* and *Trans* algorithms and for this reason the continued use of these names will remain. However it must be noted that the versions in *SOMO* are an extension to the original program that will generate different models. Therefore any reference to either *Trans* or *AtoB* will be in context of the modified program and not the original version. The aim of this section is to describe in more detail both the *Trans* and *AtoB* algorithms and the processing operators that allow the algorithms to be optimised by the user.

4.2.1 BEAD MODEL GENERATION

SOMO provides the option of generating bead models using either the *AtoB* or *Trans* algorithm. Within *AtoB*, a cubic lattice is placed over the atomic model; the centre of gravity (c.o.g) of those residues within each cube of the lattice marks the location of each bead as described in Figure 4.3 (A). The *Trans* algorithm essentially uses two beads to represent each residue (except for glycine). One bead represents the side chain of the residue (side chain bead), while the other bead represents the peptide bond between two residues (peptide bond bead) (Figure 4.4(I)). A terminal bead is also added to represent the terminal COOH group at the end of a protein chain.

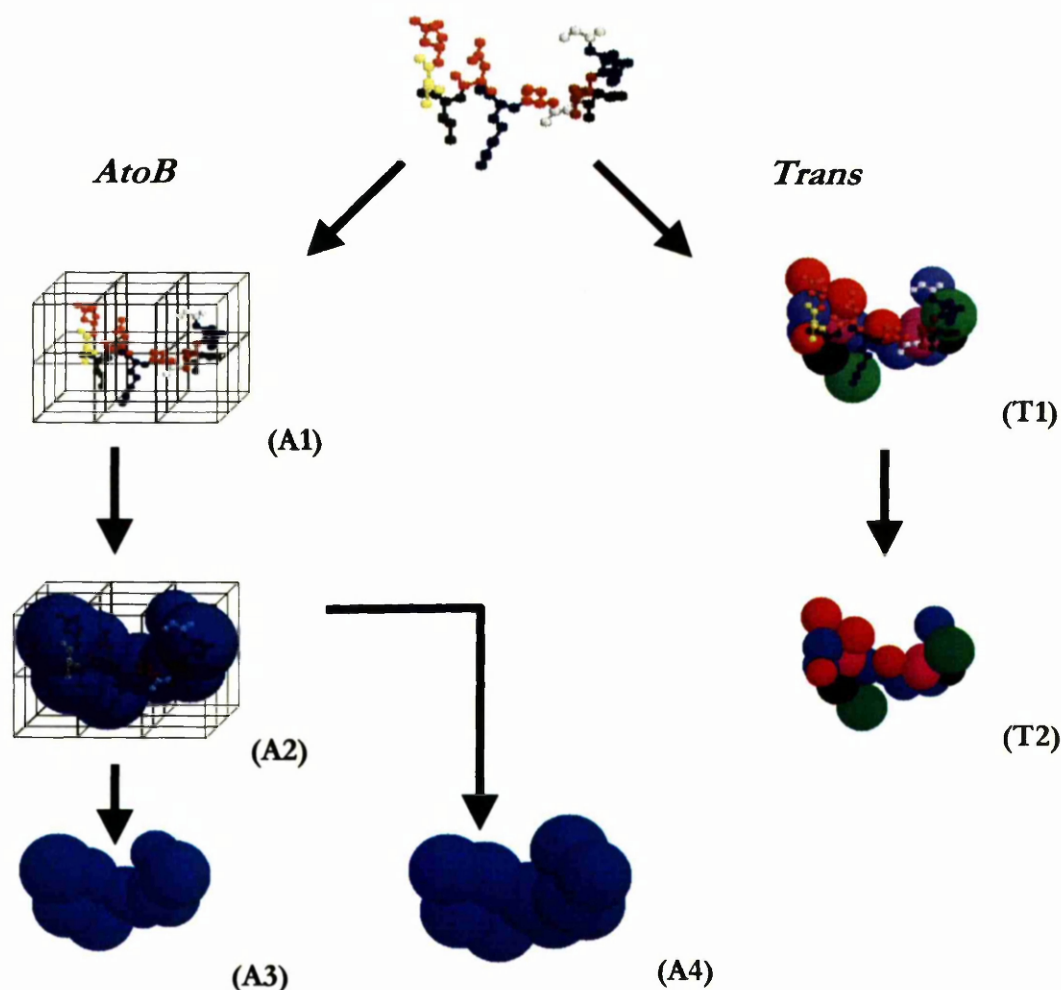


Figure 4.3 An atomic representation of a short stretch of amino-acid residues is converted to bead models according to the *AtoB* (A) or *Trans* (T) method. In *AtoB* the atomic co-ordinates are mapped onto a cubic lattice (A1). The centre of gravity (c.o.g.) for those residues in each cube marks the centre of the bead (A2). The radius of each bead can be either variable (A3) or fixed (A4). A variable radius corresponds to the sum of the anhydrous volumes for those residues in each cube, whereas a fixed radius is the average radius calculated from the molecular mass, density and total number of beads. With *Trans* two beads are placed per residue, one at the c.o.g. of the peptide bond (represented by blue beads in T1 and T2) and the other at either the c.o.g. of selected atoms from the side chain atoms or actually at the location of a single selected atom. The radius of each bead is obtained from a look-up data table described in more detail in section 4.2.5.

Two choices are offered for positioning the side chain bead, both involve selecting any number of atoms from the side chain of a residue. In the first option the side chain bead is

positioned at the c.o.g. of the selected atoms and in the second option it is located at the position of the furthest atom away from the c.o.g. of the model, for the selected atoms. Selecting the latter option will increase the “solution volume” occupied by the model (and will also result in a slightly bigger bead volume after overlap removal, see below). It is a simple attempt at better reproducing the hydrodynamically important surface architecture of the original structure, since utilising the c.o.g. for long side chains would reproduce the extension of that residue poorly.

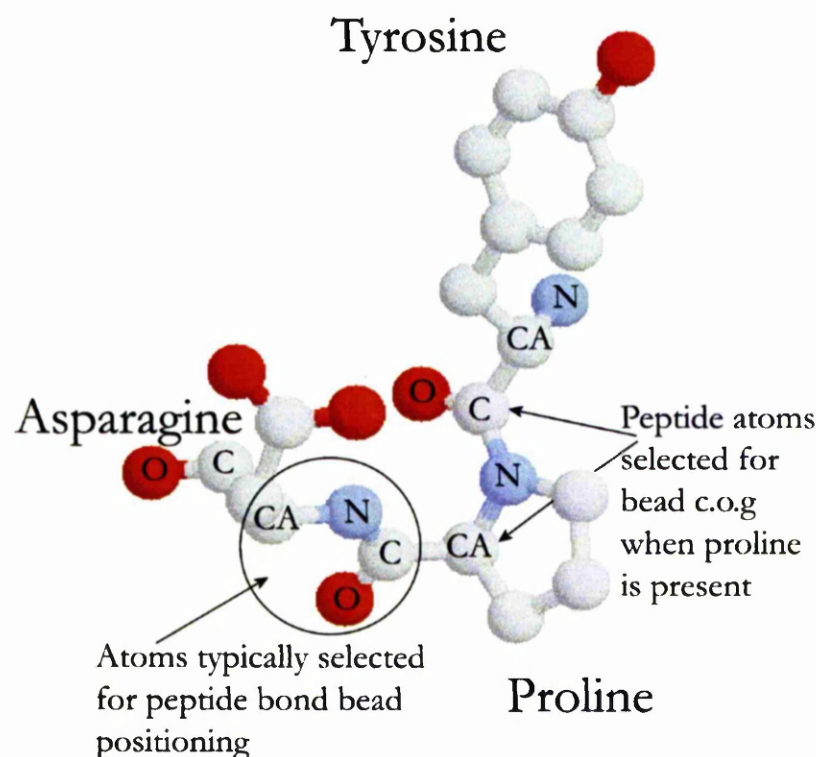


Figure 4.4 All peptide bond beads except for proline residues are placed at the c.o.g. of the four main chain atoms shared between two residues. These are the C and O of the previous residue and N and CA. For proline, only C and CA have been used because it has been found to reduce the degree of overlap otherwise encountered if all four atoms were used.

Peptide bond beads for all residues except proline are placed at the c.o.g. of the main chain atoms: C_{residue} , O_{residue} , $N_{\text{residue} + 1}$ and $CA_{\text{residue} + 1}$ (CO). Because the proline side chain is bonded to both the N and CA of the main chain atoms, its peptide bond bead has been placed at the c.o.g. of the C_{residue} and $CA_{\text{residue} + 1}$ atoms in an attempt to minimise any potential overlap with the proline side chain bead (Figure 4.4). This setting is currently coded into the program but other combinations will also be attempted to test whether they minimise overlaps even further.

One advantage of the *AtoB* algorithm is that the resolution of the model can be changed by altering the size of the cubes in the lattice (Figure 4.5). Lowering the resolution of the model will reduce bead numbers and hence speed up computations.

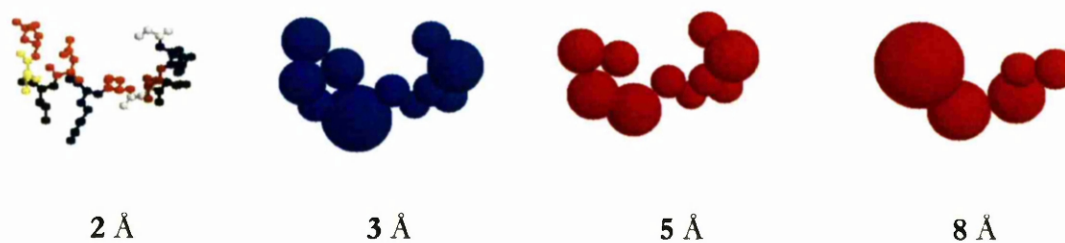


Figure 4.5 Bead models generated from atomic coordinates using the modified *AtoB* algorithm implemented in *SOMO*. The models generated at 3 Å, 5 Å, and 8 Å resolutions demonstrate how shape and bead numbers change at elevated resolutions.

Byron also highlighted two major limitations with the original *AtoB* algorithm. These were: inaccurate methods for representing the hydration of elongated molecules and an inability to remove bead overlaps which often gave erroneous intrinsic viscosity calculations for models composed of beads of differing radii when using the modified Kirkwood-Riseman approximation (Carrasco and Garcia de la Torre, 1999b). A simple solution to the latter problem was to remove overlaps between beads or to apply a modification to the interaction tensor as proposed by Zipper and Durchschlag (1998). The original *Trans* program possessed processing operators that could remove overlapping beads by the method of proportional reduction which will be described in section 4.2.2. This method has now been implemented in *SOMO* so that it can remove overlapping beads from models generated by the *AtoB* algorithm (as in Figure 4.5).

Hydration (δ) is currently represented in *AtoB* by expanding the radius of each bead uniformly. This involves translating all the beads away from the c.o.g. of the molecule by an amount dependent on the expansion factor (ϵ) (Tanford, 1961 cited in Garcia de la Torre, *et al.*, 1999).

$$\epsilon = \left(1 + \frac{\delta}{v\rho} \right)^{\frac{1}{3}} \quad \text{Equation 4.1}$$

So the translation of a bead to a new coordinate (in this particular example the x coordinate) will depend upon the original coordinates x_{orig} , the x coordinate of the c.o.g., $x_{\text{c.o.g.}}$, and ϵ as shown below (Byron, personal communication):

$$x = x_{c.o.g.} + (\mathcal{E}(x_{orig} - x_{c.o.g.}))$$
Equation 4.2

The radius of the expanded bead is determined from its original radius, r_{orig} and \mathcal{E} as shown below:

$$r = r_{orig}\mathcal{E}$$
Equation 4.3

Since the amount of translation is proportional to the distance between the original bead and the c.o.g., then those beads occupying the extremities of an elongated model will be translated further than those beads closer to its c.o.g. as illustrated by Figure 4.6, which is why this approach works better for globular rather than elongated molecules, as stated in (Byron, 1997).

An alternative method to uniform expansion that was implemented in the *Trans* program involved assigning at the beginning an hydration value to each bead, according to the theoretical values worked out by Kuntz and Kauzmann (1974). *SOMO* stores the level of hydration for each kind of residue in a data table (section 4.2.5), which can be changed without altering the actual program code. This feature offers more flexibility for adjusting the hydration levels than having the values embedded directly into the program code. This means that the hydration value for each residue can be changed according to specific environmental conditions (such as pH) that may effect it (Kuntz and Kauzmann, 1974).

The bead model generated by *Trans* accounts for hydration by enlarging each bead by an amount that would accurately reflect the number of water molecules associated with a particular bead residue. No distinction is made between exposed and buried residues, but since the former are usually the more hydrophilic ones, most of the expansion due to hydration happens at the model's surface. It must be recalled that water molecules are also often found in cavities in the protein interior (Williams *et al.*, 1994). *AtoB* on the other hand just hydrates those residues present within surface beads¹⁹ by expanding those beads by a volume proportional to the hydrated residues encapsulated in each bead. There is no specific rationale for this distinction other than that the effect of hydration tends to be more noticeable on the protein surface. However hydration of all the beads will also be investigated at a later date.

¹⁹ A surface bead in the context of models generated by *AtoB*, denotes any bead possessing a surface residue.

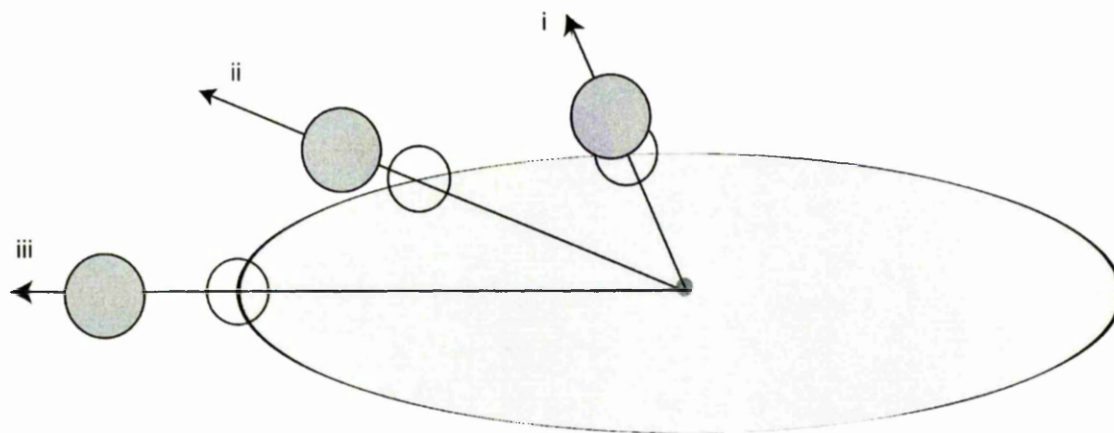


Figure 4.6 An ellipse, representing an elongated model, is used to demonstrate one of the flaws (identified by Byron, 1997) of isomorphous expansion as implemented by the *AtoB* algorithm. Isomorphous expansion is used to translate beads (depicted as transparent and shaded circles before and after translation respectively) away from the c.o.g. of the model by an amount proportional to the distance of the original bead from the c.o.g. of the model. Therefore beads situated further away (iii) from the c.o.g. of the model will be translated a greater distance than those located closer (i and ii).

4.2.2 OVERLAP REMOVAL

In order to determine hydrodynamic parameters according to the modified Oseen tensor by Garcia de la Torre and Bloomfield (1977a), beads overlapping by more than 0.001 nm are removed by reducing their radius until there is no overlap. The amount of reduction is dependent on the size of the bead and its degree of overlap with other beads. Two methods have been implemented for removing overlaps: asynchronous and synchronous bead reduction. Asynchronous bead reduction sequentially reduces the bead radii by searching for the two beads with the greatest overlap within the model. The overlap is then removed by reducing the radius of either one or both of the overlapping beads. Removing the overlap by reducing only one bead inevitably results in that bead shrinking to a very small size. If the overlap distance between the two beads is greater than the radius of the smaller bead, then the radius of the smaller bead is simply reduced to a bead size of 0.001 nm and is assigned an absent status. This means that it will not be displayed or used in certain hydrodynamic calculations but its mass will be recorded so that the overall mass of the model remains constant.

Radial reduction of two beads is done according to the equation below, where r_{A0} and r_{B0} are the radii of two beads separated by a distance $|AB|$, overlapping by an amount O . The new bead radii for both beads with overlaps removed are r_{A1} and r_{B1} (Equation 4.5).

$$O = (r_{B0} + r_{A0}) - |AB| \quad \text{Equation 4.4}$$

$$r_{B1} = r_{B0} - \left[O \left(\frac{r_{B0}}{r_{A0} + r_{B0}} \right) \right] \quad \text{Equation 4.5}$$

This is known as proportional reduction, which is an alternative to equal reduction where overlaps are removed by simply halving the overlap distance and subtracting this value from the radius of both spheres. The equal reduction method has been avoided because it can lead to the formation of very small beads as illustrated by Figure 4.7.

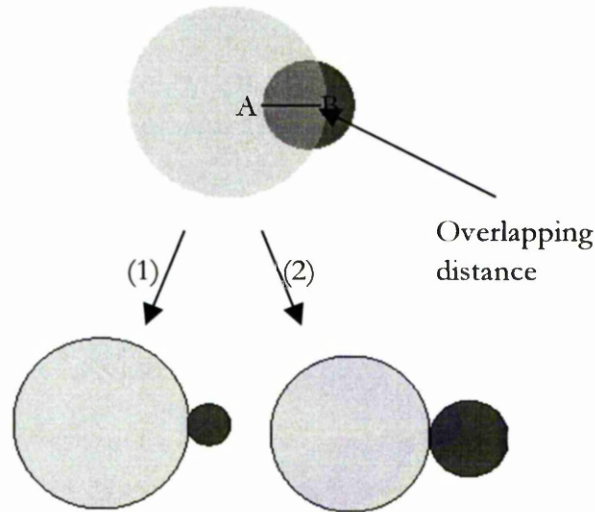


Figure 4.7 Two overlapping spheres represented by differently shaded circles illustrate the effect of proportional reduction. The bead size of the smaller, darkly shaded sphere is minimally effected by proportional reduction (2) compared with the more drastic effect of equal reduction (1).

The synchronous method of overlap removal searches for the smallest overlap between beads and then proceeds to reduce the bead radii of all those overlapping beads by the value of the smallest overlap.

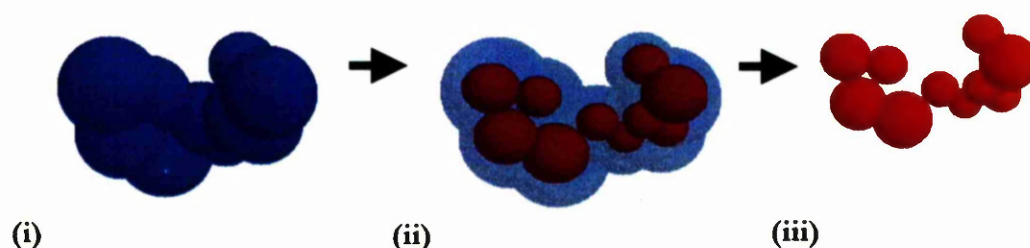


Figure 4.8 Illustrates the removal of overlaps (iii) from a model containing overlapping beads (i). The consequence of overlap removal is a significant decrease in the volume of the model and a change in its surface area and topology (ii).

The consequence of overlap removal on the model is an overall reduction in its volume (Figure 4.8), a change in its surface area and its surface contour, all of which will have a significant effect on its calculated hydrodynamic properties. The beads that contribute to the hydrodynamic properties are mainly those beads occupying the surface of the model. Therefore the overlap removal operator removes overlaps in a particular sequence so that changes made to the surface beads are minimised. The general outline of this sequence as applied to the *AtoB* algorithm, is that overlaps are first removed only from those beads at the surface using the proportional reduction method. Then overlaps between surface and buried beads and buried and buried beads are removed so that only buried bead radii are reduced in the former case and both are proportionally reduced in the latter. The reason for the different methods of reduction in the latter case is to try to minimise the changes affecting the surface beads. The simple pseudo-code for this latter procedure is as follows:

- *If surface bead overlaps buried bead then reduce only the radius of the buried bead provided its radius is large enough to support the reduction.*
- *If not then reduce the buried bead radius to 0.001 nm and exclude it from any further calculations.*
- *If two buried beads overlap then proportionally reduce both beads*

For *Trans* this process is slightly more complex because each bead is not only classified according to its surface accessibility, but also whether it is a side chain or peptide bond bead. The sequence of events for both the *Trans* and *AtoB* algorithms are described in Table 4.2. Each box indicates an action that is to be taken for the types of beads in the box. For example, step 2 of *Trans* – (surface s.c. \Leftrightarrow surface p.b. and surface p.b. \Leftrightarrow surface p.b.) indicates that if a surface side chain bead overlaps with a surface peptide bead, then only the radius of the latter is reduced because it is the only one underlined. Whereas if the overlapping beads are surface peptide beads then both are reduced as both of them are underlined.

To minimise the change in the excluded molecular volume (Paci and Velikson, 1997), the number of beads being reduced at the surface must be minimal. Beads encountered at the surface of a molecule overlap with either other surface beads or buried beads. Reducing only the buried bead from a buried-surface bead overlap is one approach being used to minimise the number of surface bead radii modified.

A further distinction can be applied to the *Trans* algorithm by differentiating surface side chain beads with peptide bond beads. Non hydrophobic surface side chains generally extend from the surface of the molecule and interact with solvent and other molecules. Because they have a crucial role in hydration, it is preferable to maintain side chain bead radii by reducing only surface-peptide bond beads for a surface side chain bead – surface peptide bond bead overlapping pair (surface s.c. \Leftrightarrow surface p.b.).

Steps	<i>AtoB</i>	<i>Trans</i>
1	<u>surface</u> \Leftrightarrow <u>surface</u>	<u>surface s.c.</u> \Leftrightarrow <u>surface s.c.</u>
2	surface \Leftrightarrow <u>buried</u> and <u>buried</u> \Leftrightarrow <u>buried</u>	surface s.c. \Leftrightarrow <u>surface p.b.</u> and surface p.b. \Leftrightarrow surface p.b.
3		surface \Leftrightarrow <u>buried</u>
4		<u>buried</u> \Leftrightarrow <u>buried</u>

Table 4.2 Overlap removal for both *AtoB* and *Trans* is a multi-step process requiring the reduction in the radius of a bead to remove overlaps. surface and buried correspond with beads either encountered at the surface or the interior of a molecule respectively. With the *Trans* algorithm, the location of the bead may have either an s.c. or p.b. extension denoting side chain and peptide bond respectively. Only those labels that are underlined are undergoing radius reduction. \Leftrightarrow refers to those bead types that are being compared with each other.

The sequence of steps used by the overlap removal operator is a first step towards minimising the number of beads that have to undergo volume reduction. Other methods that have been incorporated into *SOMO* include outward translation (section 4.2.3) and bead popping (section 4.2.4), the former being previously implemented into the *Trans* program.

4.2.3 OUTWARD TRANSLATION

Chapter 3 briefly explained how the size and position of the outermost beads of a model accounts for its frictional coefficient, ζ and molecular volume, V . The multi-step overlap removal procedure (mentioned above) was designed to minimise the number of surface bead radii being reduced. However if the molecule is more elongated, the overall surface and the volume reductions of the model may be very significant. The outward translation operator attempts to compensate for this reduction in surface bead radius by translating that bead away from the c.o.g. of the model. The extent of this translation depends on the amount of radius reduction. Figure 4.9 illustrates the procedure using a circle to represent a slice of a globular protein. Two overlapping spheres (Figure 4.9 (1)) on the outer surface of this molecule are reduced in size until there is no further overlap (Figure 4.9 (2)). This results in a reduction in the circumference of the molecule, denoted by the shaded area of the circle. However, applying the outward translation operator (Figure 4.9 (3)) results in both overlapping beads translating in a direction so that the outer circumference of the beads coincides with the outer circumference of the molecule, thus maintaining the original surface topology.

The outward translation currently uses the method illustrated in Figure 4.9 (4) along with the equations outlined below to determine the translated coordinates (x_3 , y_3 and z_3) of the bead. The distance between the c.o.g. of the model and bead is r , which can be determined from:

$$r = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad \text{Equation 4.6}$$

The relation between r and the translation distance δr is (for the x coordinate):

$$\frac{(x_3 - x_2)}{(x_2 - x_1)} = \frac{\delta r}{r} \quad \text{Equation 4.7}$$

and similarly for the y and z coordinates. So the value for x_3 is:

$$x_3 = \left[\frac{(\delta r(x_2 - x_1))}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}} \right] + x_2 \quad \text{Equation 4.8}$$

This also applies in the y and z coordinates. Translated coordinates can also be determined by vector geometry using the vector library. This had previously been implemented into the program and gave the same values as the current implementation.

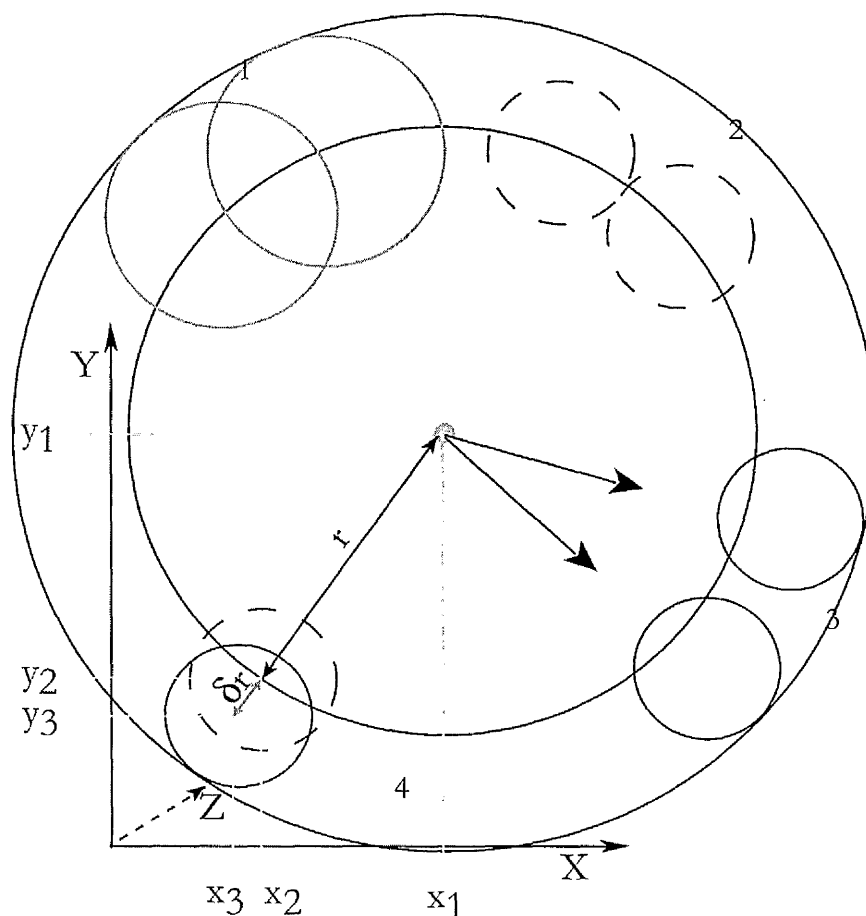


Figure 4.9 The outer circle describes a hypothetical spherical molecule. In (1) 2 beads of the same diameter with a 50% overlap reside on the exterior of the molecule. With normal overlap removal both beads are reduced equally in size until there is no overlap (2). This also results in a decrease in the circumference of the molecule. To compensate for this “shrinking effect”, bead size reduction is followed by translating that bead away from the molecules’ centre of gravity (3). (4) The new coordinates for the translated bead (x_3 , y_3 and z_3) are determined from the original bead coordinates (x_2 , y_2 and z_2), the coordinates for the c.o.g. of the model (x_1 , y_1 and z_1), the distance between the c.o.g. and the bead (r) and the distance to be translated (δr).

The current implementation could be further improved in at least two aspects:

- Translating beads in a direction away from the surface of the model instead of its c.o.g..
- Reducing the beads by an amount that would minimise the gaps between two previously overlapping beads.

A bead translated in a direction perpendicular to the surface of the model would reflect more accurately the expansion of the model as a result of hydration. The problem with the current

implementation is that a bead translated away from the centre of gravity of a non spherical model still results in an inaccurate representation of the expansion process.

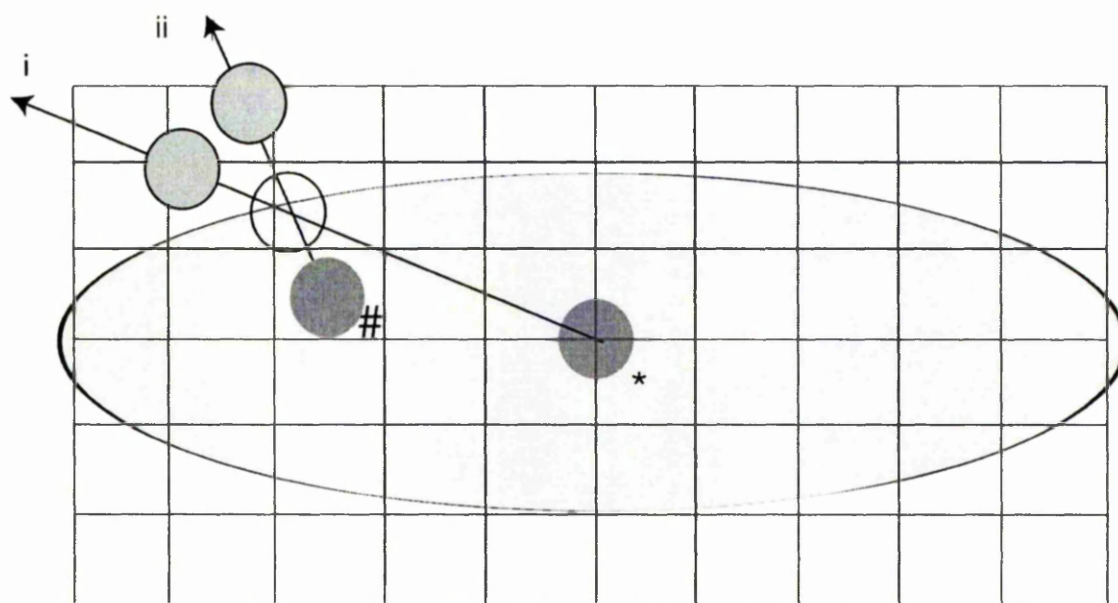


Figure 4.10 The difference between (i) translating a bead from the centre of gravity of the molecule (*) and (ii) translating it from a region closer to the site of translation (#).

Take, for example, the elliptical shape in Figure 4.10, which describes a slice taken from a hypothetical molecule. The current implementation as in Figure 4.10 (i) is the translation of a bead away from the centre of gravity (*) of the molecule. The alternative to this is a translation that is more perpendicular to the surface of the molecule as in Figure 4.10(ii), so that the model is expanded more evenly. One approach for doing this involves applying a cubic grid to the whole molecule and then determining the c.o.g. for those cubes containing only buried beads that are closest to the bead being translated. For instance, in Figure 4.10, a cubic mesh has been applied to the model. The c.o.g. for the buried cube closest to the bead being translated (darkly shaded circle with (#)) is used as the new reference point for translation. The principle of this technique is to simply move the reference point closer to the surface of the model, in particular to the actual bead being translated. This should cause an even, more expanded outer surface for regions not lying on the equator and at the central axis of the model.

Although this is a relatively simple method to implement its major drawback can be seen when returning to a spherical example as in Figure 4.11 (i). In this example, the buried beads closest to the translating bead occupy squares A and B. The first problem is that the direction

of translation of the surface bead can change significantly depending on its proximity to beads in cube A or B. The second problem is that translation from either of these locations is significantly far from being perpendicular to the surface of the model.

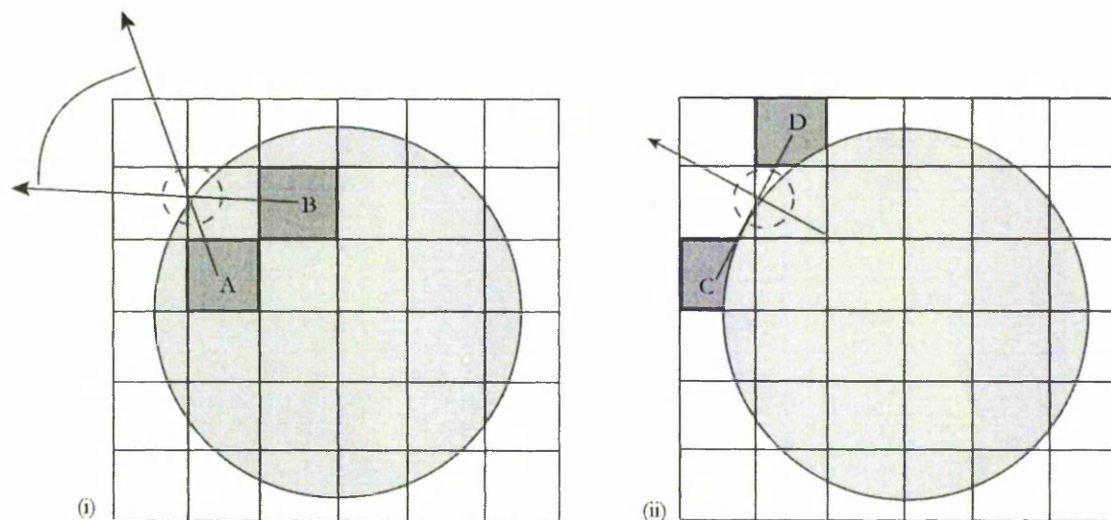


Figure 4.11 (i) The problem of translating a bead from a reference point close to the point of translation is illustrated here. A and B are two points on the grid representing regions close to a bead (represented by a dashed circle) that is to be translated. The locations of these reference points highlight the difficulty of translating beads perpendicular to the surface of the model (shaded circle). One solution to this problem (ii) is to identify reference points (C and D) which lie parallel to the surface of the model. The bead can then be translated perpendicular to this plane.

An alternative method as shown in Figure 4.11 (ii) is to determine the plane of the surface of the model local to the bead to be translated. In this example, the plane is defined by two points taken from surface cubes neighbouring the cube containing the translating bead. Clearly this example is two dimensional, but the same principles apply in three dimensions but with the selection of additional points in the third dimension. Translation would then be made perpendicular to the surface of this plane directed away from the bulk of the model. Whether this method would reflect more accurately the volume and surface topology of the molecule is yet to be tested. But what is clear is that translation away from the c.o.g. of the model and translation away from the surface would each have different effects on the shape of the model.

The second improvement is the removal of gaps between two previously overlapping beads. This can be achieved using either a analytical treatment or a computationally intensive method. For the analytical treatment, consider two overlapping beads (centred at A_0 and B_0) at the surface of the model as illustrated in Figure 4.12.

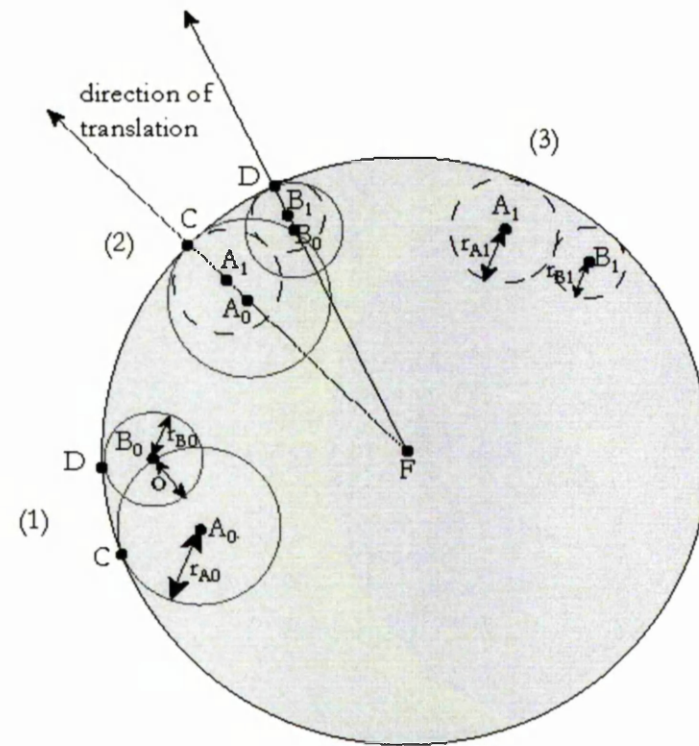


Figure 4.12 Illustrates a mechanism for overlap removal that eliminates gaps which would have otherwise been present in the original outward translation procedure. In this picture, a shaded circle depicting a slice through a spherical model whose c.o.g is marked by F encloses two spheres (1) with radius r_{A0} and r_{B0} overlapping by O. The centre of both spheres is A_0 and B_0 and their furthest point away from the c.o.g is C and D respectively. Translation and proportional reduction of both beads so that they no longer overlap will give new bead positions of A_1 and B_1 with radii r_{A1} and r_{B1} respectively.

To remove these overlaps, both beads have to undergo radial reduction from r_{A0} to r_{A1} and r_{B0} to r_{B1} . To ensure that the surface contacts C and D are maintained, the following conditions have to be satisfied. (i) Both beads are translated away from the c.o.g of the model, from their initial starting point of A_0 and B_0 to A_1 and B_1 respectively. (ii) The distance $|FC|$ and $|FD|$ remain unchanged during the translation and radial reduction process.

$$|FC| = |FA_0| + r_{A0} = |FA_1| + r_{A1} \quad \text{Equation 4.9}$$

and

$$|FD| = |FB_0| + r_{B0} = |FB_1| + r_{B1} \quad \text{Equation 4.10}$$

(iii) The distance between A_1 and B_1 equals $r_{A1} + r_{B1}$.

$$|A_1B_1| = r_{A1} + r_{B1} \quad \text{Equation 4.11}$$

(iv) That both beads undergo proportional reduction.

$$\frac{r_{A0}}{r_{B0}} = \frac{r_{A1}}{r_{B1}} \quad \text{Equation 4.12}$$

The alternative, computational method for removing the gaps between two outwardly translating beads involves reiterating the overlap removal by increments that are a fraction of the total overlap. This would continue until the overlap between the beads is removed. The size of the gaps between translating beads will be proportional to the size of the fraction overlaps being removed. This method although computationally intensive, will reduce bead radius proportionally. Both computational and analytical solutions have yet to be implemented into the *SOMO* program.

4.2.4 BEAD POPPING

The bead popping operator was written specifically for the *Trans* algorithm as a precautionary measure to avoid very small beads as a result of removing the overlaps between two largely overlapping beads. Figure 4.13 illustrates this with a simple stretch of amino-acid residues, bead generation and the application of the bead popping operator.

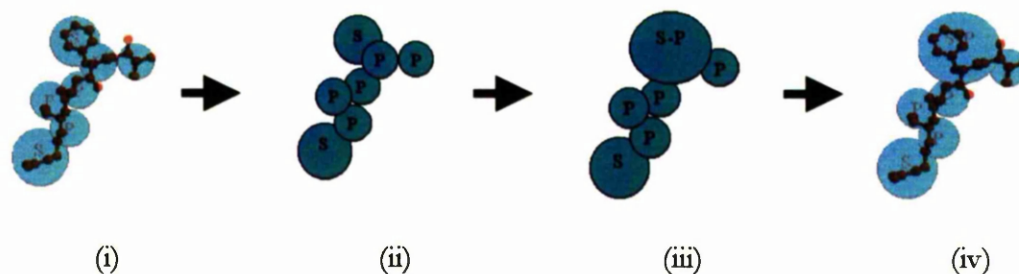


Figure 4.13 Sequence of steps starting from the generation of a bead model from a short stretch of residues (i) using the *Trans* algorithm. Each bead is represented (ii) either as a peptide (p) or a side chain (s) bead. An S and P bead have fused (iii) because they overlapped more than a specified value. Fusion between two beads should have little effect on the fit of the original atomic coordinates to the bead model (iv).

The decision to fuse two beads is dependent on determining the percentage of overlap between two beads (Figure 4.14(i)). This requires knowledge of the total volume V of the overlap which is $v_d + v_l$, where v_d and v_l are the volumes of the caps of the small dark shaded sphere and larger lighter shaded sphere respectively (Figure 4.14(ii)). The volume of one of these caps is conceptually equivalent to the sum of areas for all the slices taken along the x axis of the cap. For example, the area of each slice for the lighter shaded cap is $a = \pi f(x)^2$, where $f(x)$ is the radius of the slice at a point along p_l . This area will vary depending on where it is taken along the x axis. At $x = r$, (x being the distance along the x axis starting from the origin of the sphere) the area will be zero and maximal at $x = r - p_l$ (p_l being the distance along the x axis of the lighter cap).

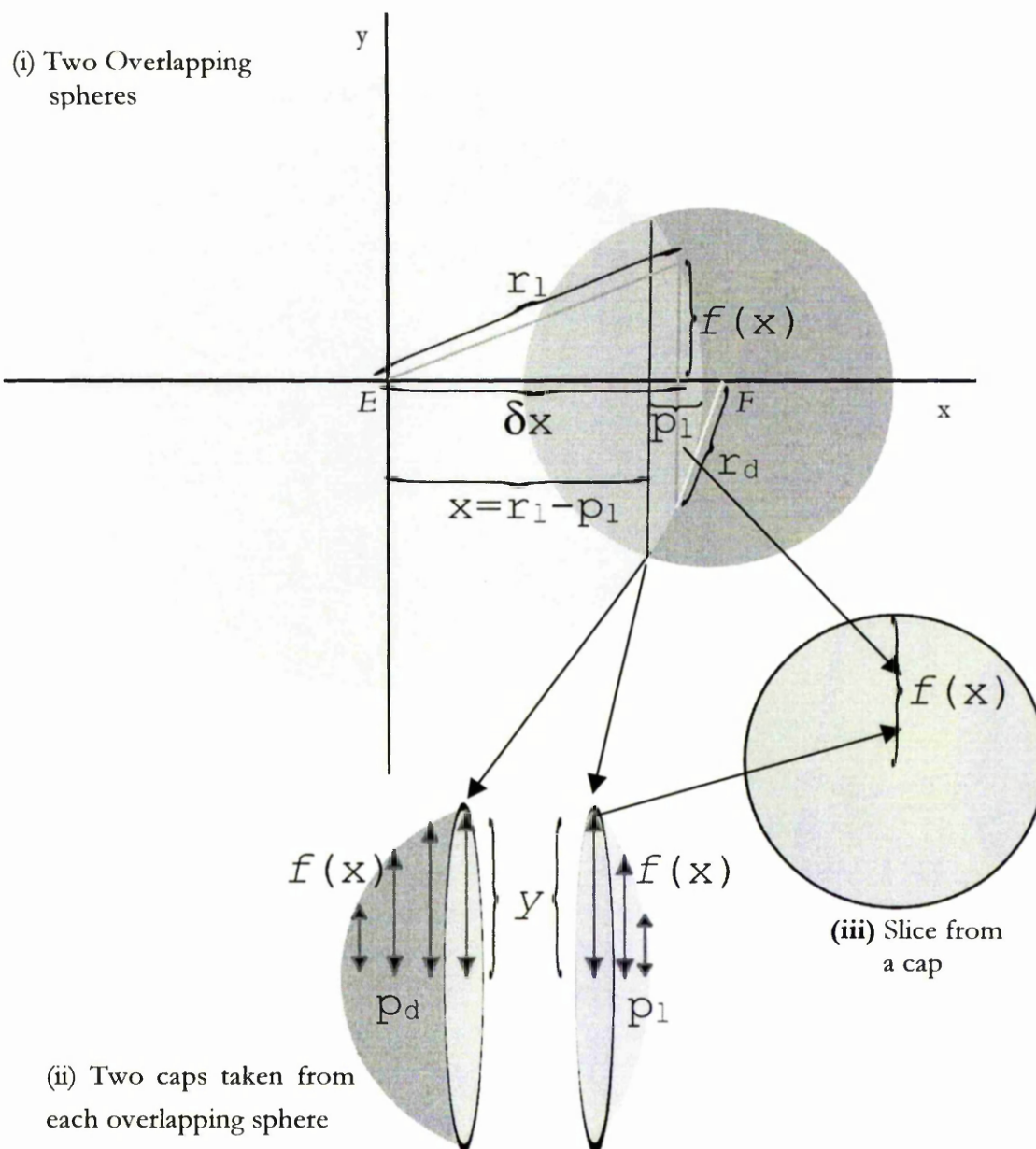


Figure 4.14 Two spheres (i) represented as differently shaded circles; light (l) and dark (d) with a radius r_l and r_d and their centre points E and F , corresponding to the light and dark spheres respectively. Both overlap by $p_l + p_d$. The total volume of overlap V is the combined volume $v_l + v_d$ for the overlapping caps (ii) of both spheres. The volume of each cap is the sum of the areas of all the slices (iii) from the cap. The radius of each slice ($f(x)$ – depicted by the double ended arrows in (ii)) will vary depending on where it is located along the p axis.

If a slice has a radius $f(x)$ at position x , (Figure 4.14(iii)) then its area will be:

$$A = \pi[f(x)]^2 \quad \text{Equation 4.13}$$

$f(x)$ is:

$$[f(x)]^2 = r^2 - x^2 \quad \text{Equation 4.14}$$

The volume is the sum of the areas of all the slices from both caps ($v_l + v_d$) lying between E and F along the x axis. The volume for a single cap, for example the lighter cap is effectively the area of all the slices between $r-p$ and r :

$$v_l = \int_{r_l-p_l}^{r_l} \pi[f(x)]^2 \delta x \quad \text{Equation 4.15}$$

Substituting Equation 4.14 into equation 4.15

$$v_l = \int_{r_l-p_l}^{r_l} \pi[r_l^2 - x^2] \delta x \quad \text{Equation 4.16}$$

Which then gives:

$$v_l = \pi \left[r_l^2 x - \frac{x^3}{3} \right]_{r_l-p_l}^{r_l} \quad \text{Equation 4.17}$$

Entering the values for $x=r$ and $x=r-p$

$$v_l = \pi \left[\left(r_l^3 - \frac{r_l^3}{3} \right) - \left(r_l^2(r_l - p_l) - \frac{(r_l - p_l)^3}{3} \right) \right]$$

$$v_l = \pi \left[\frac{r_l p_l^2 - p_l^3}{3} \right] \quad \text{Equation 4.18}$$

So the general expression for the volume of the lighter cap is:

$$v_l = \frac{1}{3} \pi p_l^2 (3r_l - p_l) \quad \text{Equation 4.19}$$

and the total volume of the overlap is:

$$V = v_d + v_l = \frac{1}{3} \pi [p_d^2 (3r_d - p_d) + p_l^2 (3r_l - p_l)] \quad \text{Equation 4.20}$$

Both r_d and r_l are known which leaves only p_l and p_d to be determined. The distance between E and F is:

$$EF = (r_d - p_l) + (r_l - p_d) \quad \text{Equation 4.21}$$

r_l and r_d are:

$$r_l^2 = (r_l - p_d)^2 + y^2 \quad \text{Equation 4.22}$$

$$r_d^2 = (r_d - p_l)^2 + y^2 \quad \text{Equation 4.23}$$

To get p_d Equation 4.21 is re-arranged to make $r_l - p_d$ the subject. This is then substituted into Equation 4.22 to give:

$$r_l^2 = (EF - (r_d - p_l))^2 + y^2 \quad \text{Equation 4.24}$$

Subtracting Equation 4.23 from Equation 4.22 gives:

$$\begin{aligned} r_l^2 - r_d^2 &= (EF - (r_d - p_l))^2 + y^2 - y^2 - (r_d - p_l)^2 \\ r_l^2 - r_d^2 &= EF^2 - 2EF(r_d - p_l) + (r_d - p_l)^2 - (r_d - p_l)^2 \\ r_l^2 - r_d^2 &= EF^2 - 2EFr_d + 2EFp_l \end{aligned} \quad \text{Equation 4.25}$$

Re-arranging the above equation to get p_l :

$$p_l = \frac{r_l^2 - r_d^2 - EF^2}{2EF} + r_d \quad \text{Equation 4.26}$$

Substituting the expression for p_l into Equation 4.21 will give p_d :

$$\begin{aligned} EF &= (r_d - p_d) + \left(r_l - \left(\frac{r_l^2 - r_d^2 - EF^2}{2EF} + r_d \right) \right) \\ p_d &= r_l - EF + \frac{EF^2 - r_l^2 + r_d^2}{2EF} \end{aligned} \quad \text{Equation 4.27}$$

So the total overlap V can be obtained from substituting equation 4.26 and equation 4.27 into Equation 4.20. The percentage overlap of a bead is $(V/TVB) \times 100$, where TVB is the total volume of the bead.

4.2.5 LOOK-UP DATA TABLE

The processing operators discussed so far provide methods for altering the size and position of groups of beads but limited action is allowed for beads representing particular types of residues. But because each residue possesses unique characteristics such as hydrophobicity, charge and polarity which affects the overall structural arrangement of the molecule (packing) and its interaction with the solvent environment (hydration) (Thanki *et al.*, 1988), it was necessary to provide a facility that would take these features into consideration.

The current implementation offers a look up table (Figure 4.15) which lists the radii (and hence volume) assigned to side chain, peptide bond and terminal beads; the number of water molecules associated with each component (and hence its increase in volume); its colour and its position with respect to the side chain atoms of that residue. By providing this information in the form of a look-up data table, specific changes can be made to the properties of a bead representing a particular side chain residue. This is very useful especially for studying how the hydrodynamic properties of a molecule changes with the hydration of specific residues.

The table was designed upon the requirements of the *Trans* routine but some of the information from the table such as the water volume, M_r for each residue and its level of hydration, can also be used by *AtoB*. The *AtoB* selection in *SOMO* also has an alternative option for using data values taken from the original program. This not only includes the M_r and hydration values for each residue but also information about its SAS length relevant for Debye sphere modelling.

The radii for each residue side chain and peptide bond bead was calculated from the volumes of each residue as determined by Tsai *et al.* (1999) using Voronoi polyhedra. The peptide bond bead was assumed to share the same volume as the glycine residue, so that all other residue volumes had the value of 64.9 Å³ subtracted from it. These volumes were derived specifically for *Trans* but can also be used for *AtoB*. Or alternatively, the *AtoB* algorithm also allows volumes to be calculated directly from the partial specific volume, \bar{v} of each residue as in the original *AtoB* program. Values for \bar{v} were obtained from Perkins (1986) who used crystallographic data to determine the volumes for each amino acid residue. Although the correlation between residue volumes determined by this method and from densimetry were considered good, the volumes were generally larger in the densimetry approach. This difference was attributed to hydration effects that were *seen* by densitometry and not by the crystallographic approach. Therefore \bar{v} based on sequence composition can be regarded as the anhydrous volume of those residues, which means that bead models have to account for hydration independently i.e. by assigning a set number of water molecules to each residue (or just the surface residues in the case of *AtoB*). The number of water molecules was obtained

directly from Kuntz and Kauzmann (1974), while the volume of the water molecule was obtained from Gerstein and Chothia (1996). The last column declares whether the beads should be placed at the c.o.g. of the atoms selected in the first five columns (0) or at the position of the atom sited the furthest away from the c.o.g. of the model.

22	24.041										
CB	XX	XX	XX	XX	71.10	15	9	0	1.816	ALA	1
SG	XX	XX	XX	XX	103.20	47	2	0	2.183	CYS	1
OD1	OD2	XX	XX	XX	115.10	58	3	5	2.321	ASP	0
OE1	OE2	XX	XX	XX	129.10	72	3	6	2.643	GLU	0
CD2	CE1	XX	XX	XX	147.20	91	9	0	3.118	PHE	0
XX	XX	XX	XX	XX	57.10	0	0	0	0.000	GLY	2
ND1	NE2	XX	XX	XX	137.20	73	4	3	2.832	HIS	0
CB	CG1	CG2	CD1	XX	113.20	57	9	0	2.870	ILE	0
CE	NZ	XX	XX	XX	128.20	73	4	3	2.902	LYS	0
CG	XX	XX	XX	XX	113.20	57	9	0	2.871	LEU	1
SD	XX	XX	XX	XX	131.20	75	9	1	2.899	MET	1
OD1	ND2	XX	XX	XX	114.10	58	2	1	2.426	ASN	0
CB	CD	XX	XX	XX	97.10	41	9	1	2.401	PRO	0
OE1	NE2	XX	XX	XX	128.10	72	2	1	2.722	GLN	0
NH1	NH2	XX	XX	XX	156.20	101	4	2	3.135	ARG	0
OG	XX	XX	XX	XX	87.10	31	2	1	1.938	SER	1
OG1	CG2	XX	XX	XX	101.10	45	2	1	2.382	THR	0
CB	CG1	CG2	XX	XX	99.10	43	9	0	2.606	VAL	0
CD1	CE3	CZ3	XX	XX	186.20	130	9	1	3.391	TRP	0
CE1	CE2	XX	XX	XX	163.20	107	2	2	3.159	TYR	0
CA	C	O	N	XX	56	56	2	1	2.493	PEP	0
OXT	XX	XX	XX	XX	58	58	8	5	2.561	TER	1

Figure 4.15 Example look-up table. The two values in the first line (highlighted in yellow) correspond to the number of lines in the look up table that is to be read in by *SOMO*, and the value for the volume of water (in Å) associated with each residue component (Gerstein and Chothia, 1996). The remaining lines, whose values are colour coded correspond to: the **atom names** (first five columns) that are to be included for determining the position in which each bead is to be placed; the molecular weight, M_r of each **residue** and for its **side chain component** (except in the last two lines where it corresponds to the main chain and OXT terminal residue); the **colour** value assigned to a bead representing a particular residue; the number of **water molecules** associated with each residue component, its **radius** (Tsai *et al.*, 1999), **name** and the type of method used to place the bead. A value of 1 will place the bead at the position of the furthest atom away from the c.o.g. of the model while a value of 0 will place the bead at the c.o.g. of the list of atoms. Only in the case of glycine will neither routine be called and will therefore have a value of 2.

4.3 GRAPHICAL INTERFACES

The last decade has witnessed the development of a considerable number of molecular modelling packages. This can be partly attributed to the lower cost of computing facilities and improvements to the user interface, both of which have made these programs more accessible and user friendly. The widespread use of computers in recent years has resulted in the design of more intuitive GUT's (graphical user interface). This has partly been achieved by standardising the interface design so that it obeys a format that can be recognised by regular users. Currently desktop interfaces utilise desktop icons and pull down menus as a means of providing intuitive guides for using software. Similar designs are used for software possessing a graphical interface.

SOMO was developed on a Silicon Graphics O₂ workstation running the UNIX operating system (IRIX 6.3). UNIX, renowned for its robustness and reliability, has always been popular with the academic and industrial sector for developing and running more demanding applications. Although the requirements at the outset of this project could be met by the more widely available operating systems such as Windows and the Mac operating system, UNIX was chosen because of its reliability and popularity with other software developers.

SOMO currently runs independently of a graphical interface, which means that it is more portable to a variety of different operating systems. It generates an output file that can be read by the intermediary program *RAZ*, which converts SOMO output files to two script files that can be read by *Rasmol*. SOMO can also be accessed by *MOLed*²⁰ for displaying the models. *MOLed* is currently unfinished and as such a variety of other programs (specifically shareware) have also been considered for displaying bead models. Table 4.3 reviews some of the more general features available from five different graphics programs that have been used for displaying bead models. A brief explanation for each of the features is provided below.

²⁰ The latest version of SOMO does not have the link with *MOLed* enabled.

Features	<i>Mac- Beads</i>	<i>VMD</i>	<i>MOLed</i>	<i>Rasmol</i>	<i>Swiss PDBviewer</i>
Version	1.0.1	1.6	1	2.6	3.5
Availability of source code		√	√ ♠	√	
Operating systems:					
UNIX		√	√	√	√
Mac OS (Macintosh)				√	√
MS Windows (PCs)	√	√		√	√
Model manipulation					
With mouse	√ ♥	√	√	√	√
With text	√	√		√	
Saving changes	√				√
Graphics					
Rendering options		√		√	√
Colour selection		√	√	√	√
Stereo viewing		√	√	√	√
Display					
Multiple models		√	√		√
Residue/ bead selection		√	√	√ ♣	√
Animate models		√	√		
Labelling		√	√	√	√
Additional features		√			√

Table 4.3 *SwissPdbViewer* has recently changed the name of its program to *Deep View*. ♠ The source code was made available because *MOLed* was being developed at the University of Glasgow, but this has not been made available to the public domain yet. ♥ Does not allow rotational manipulation of the model. ♣ Animations can be run through CHIME (Columbia Hypermedia Immersion Environment) (Dossick and Kaiser, 1999).

4.3.1 AVAILABILITY OF SOURCE CODE

The source or program code offers access to the actual procedures for displaying the models. Its availability was considered important at the initial stages of this project so that the bead modelling procedure could be seamlessly integrated into the code. Although an integrated interface became less of a priority, the availability of the code in the public domain meant that other developers could contribute to improving the program even further.

4.3.2 TYPES OF OPERATING SYSTEMS

Operating systems essentially coordinate the transfer of information between different hardware devices and manage the programs running on the computer. The existence of different operating systems is partly due to variations in the architecture of some hardware devices and differences in preference for the way in which applications are managed. If a program is to be accessible to a wider audience then it is important that it can be run on most of the commonly used operating systems. These include the Irix operating system for Silicon Graphics; Solaris for Sun Microsystems; Linux and Windows for IBM-like PCs; Mac OS for the Macintosh PCs and MS Windows for the IBM-like PCs. Except for *MacBeads*²¹, which only ran on the Mac OS, the other packages could be run on most other operating systems.

4.3.3 BASIC MANIPULATIONS OF MODELS

For any type of computer modelling, it is essential to provide features for manipulating the model. Routine manipulation commands include translation, rotation and scaling capabilities. The ability to manipulate individual beads or units is very beneficial especially for bead modelling as it provides greater flexibility for testing different structural conformations.

Currently the most popular if not the most convenient method of model manipulation is with a mouse or track ball but alternative methods include text based methods. The advantage of text based methods is that it can be quicker than calling up a window box and many experienced users prefer to interface via a keyboard, rather than a mouse.

The provision to manipulate or modify a model and save those changes to a coordinate file is essential when adjusting a model so that it corresponds with experimentally determined biophysical parameters. This feature currently only applies to *MacBeads* and *SwissPDB-Viewer* (Guex and Peitsch, 1997).

²¹ a program that was written by Dr Dan Thomas then at the National Centre for Macromolecular Hydrodynamics (NCMH), University of Leicester, UK for displaying and manipulating bead models within the Macintosh operating system

4.3.4 GRAPHICS AND VISUALISATION

Visualising the model can be enhanced by a variety of different rendering methods and options for assigning colours to different components of the model makes analysing different aspects of the molecular structure easier. With stereo viewing, models are placed next to each other at slightly different orientations which when viewed cross-eyed gives a 3-D impression of the molecule, which can make manipulating the model easier.

4.3.5 DISPLAY FEATURES

One very important requirement in developing a modelling package is the ability to display multiple models. This has at least two consequences:

The interaction between different molecules can be studied.

Two identical molecules can be compared under different representations.

Studying the interaction between different molecules is clearly one of the objectives for this package but also if one is to identify key residues in a bead it is beneficial to be able to display a high resolution model and see which residues correspond to which beads. Figure 4.16 is an example taken from *MOLed* showing both a bead and ribbon model of pneumolysin. A selection made on the ribbon model, which is highlighted in red, is also highlighted in red on the bead model. This ability to relate two different representations can be used to identify residues that may be involved in binding two interacting molecules (see Chapter 3). Alternatively, labels can also be used to identify key residues in beads. Although multiple models can be displayed by some packages, the correlation between bead and atomic representation has only been implemented in *MOLed*.

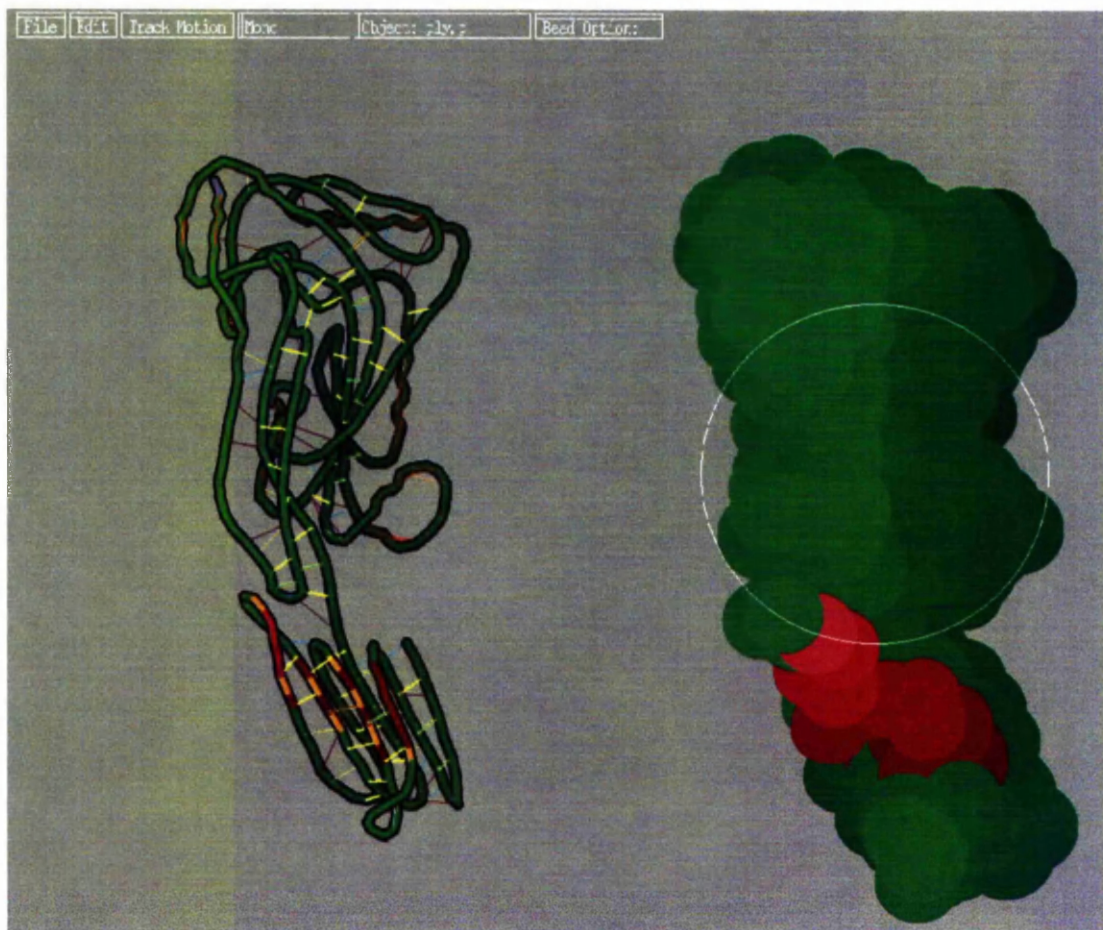


Figure 4.16 Chain and bead representation of a molecule (homology model for pneumolysin). Beads selected in the bead model are shaded in pink. They correspond with the pink residues on the chain model.

The ability to animate models i.e. moving parts or whole models along a trajectory is currently used in molecular simulations. Bead models for solution modelling have not used this feature but it could potentially be utilised by bead ‘type’ models used in folding simulations.

4.3.6 ADDITIONAL FEATURES

The features that have been discussed so far are relatively simple but are essential for any modelling procedure. However some packages such as VMD and SwissPDBViewer provide additional features such as scripting facilities. This can be used to personalise a program to perform tasks not offered in the downloaded version without the need to change the source code.

CHAPTER 5

TESTING THE ALGORITHMS ON WELL CHARACTERISED PROTEINS USING HYDRODYNAMIC THEORY

5.1 INTRODUCTION

The bead modelling studies listed in Garcia de la Torre *et al.* (2000) have been successfully applied to a variety of different experimental systems. While even poor molecular representations (where models are only represented by a few beads) are useful to gauge the solution properties of a molecule, the benefit of a more accurate description is that the calculated properties are likely to be closer to experimental values. The only limitation with either approach is the availability of information for generating the starting model.

It used to be the case that high resolution models were much scarcer for large macromolecules than electron microscopy images. However this has changed with advances in molecular techniques and the provision of automated procedures for screening protein crystals (Abola *et al.*, 2000). The growth in structural information derived from experimental data and in genome sequences has led to the development of a wide array of structure prediction algorithms (Jones, 2000), making the development of computational tools for generating medium resolution bead models from high resolution atomic structures more attractive.

Generating bead models from atomic data is essentially a process of reducing the information needed to describe the position of each atom in a macromolecule. The emphasis for hydrodynamic modelling is different from Debye sphere models. In order to conserve the overall shape of the hydrodynamic model, each bead is placed so that it reproduces the topological surface of the molecule. Whereas Debye sphere models attempt to reproduce the overall volume of the molecule and the mass distribution inside it. To do this, the volume of each bead is adjusted (with or without the inclusion of bound water) so that it reproduces the collective volume of all the atoms that it represents. Each bead is placed at the centre of gravity of a group of atoms that constitutes a scattering element.

Most bead algorithms mentioned in the previous chapter (Table 4.1) that utilise atomic data to generate the models work by grouping atoms into a single bead. This does not apply for shell models; instead a layer of small beads is placed on the outside of the macromolecule, so that the surface of the model could compare favourably with the surface rugosity of the molecule. For hydrodynamic calculations the interior of a shell model is left void because it does not contribute to the friction. In order to perform SAS calculations, the model can be filled with

equivalent sized beads so that the mass distribution from individual scattering elements is also represented. This is the basis for the *HYDROPRO* algorithm (Garcia de la Torre *et al.*, 2000), which generates both shell and 'shell filled' models. While the advantage of shell models is a detailed surface description, the number of beads that are included in the hydrodynamic calculations may compromise its speed. Instead *SOMO*, using the bead processing operators described in the previous chapter, can be used to generate medium ($> 6 \text{ \AA}$) to low resolution models.

The bead processing operators in *SOMO* can be used in a limited number of arrangements to generate a variety of different models. The algorithms that are currently working are based on grouping many atoms into a single bead. In the case of *Trans*, each bead represents either a side or a main chain (and a single terminal bead that is treated by the processing operators as if it was a main chain bead). The *AtoB* (Byron, 1997) algorithm groups atoms according to size of the grid placed over the atomic model. Using both algorithms, a variety of different operators were applied to crystallographic models of four different proteins: ribonuclease A (RNase), bovine pancreatic trypsin inhibitor (BPTI), lysozyme and citrate synthase (CS) that have already been extensively characterised. Several models were generated for the first three proteins to simulate either different hydration conditions or by including smaller numbers of beads by using either a higher grid spacing (5 \AA as opposed to 1 \AA) in the case of *AtoB* or setting a lower overlap threshold for popping in the case of *Trans*. The results from the theoretical predictions based on hydrodynamic theory were compared with experimental values obtained from solution techniques. A further test was also performed on all four models to determine how the elimination of those beads mostly present in the interior of the molecule would affect the final parameters.

5.2 METHOD

The following four computer programs were used to generate, display and calculate the hydrodynamic parameters of the bead models: *ASA*, *SOMO*, *RAZ* (which allows the visualisation of bead models by *Rasmol* (Sayle and Milner-White, 1995)), and *SUPC* (both *SUPC* and *ASA* are mentioned in Spotorno *et al.*(1997)). The programs are available on request from either Dr M. Rocco (rocco@ermes.cba.unige.it) or Dr O. Byron (o.byron@bio.gla.ac.uk) and can also be found on the enclosed CD.

The crystallographic models for each protein were downloaded from the EBI mirror of the PDB (Berman *et al.*, 2000). The accession codes are: BPTI - 4PTI, (Marquart *et al.*, 1983);

RNase – 8RAT (Tilton *et al.*, 1992); lysozyme - 1HEL (Wilson *et al.*, 1993) and CS – 1CTS (Remington *et al.*, 1982). The hydrodynamic data for lysozyme, BPTI and RNase were gathered from the original sources whenever possible, and care was taken to check that all the values had been reduced to standard conditions, e.g. water at 20°C. When this was not the case, the reduction was performed from the experimental conditions using standard formulas (Cantor and Schimmel, 1980) and tables (McRorie and Voelker, 1993). The values for CS were taken from Zipper and Durschlag (2000). The combination of processing operators used to generate each model can be found in the results and discussion section of this chapter.

5.2.1 ASA

Having acquired the appropriate PDB files, it is necessary to tag each residue as being either a surface side chain or main chain residue, or a buried residue. Arbitrary numerical codes are assigned by the program to each residue component i.e. 1 for surface side chain, 6 for surface main chain, and 10 for buried residues. The program is based on Lee and Richards' (1971) rolling sphere algorithm to determine surface accessibility. The steps required to run ASA are listed below.

Prepare the PDB file: ASA only processes information specified in the ATOM field of the PDB file, terminating at the TER field. Therefore it is preferable (but usually not necessary) to remove additional information prior to running ASA, while it is mandatory to remove any intermediate TER field for proteins containing multiple chains, otherwise *ASA* will terminate execution prematurely.

Running the program

- Launch ASAB1 at the command line.
- Supply information for the size of the rolling sphere probe as requested by the program. A typical run would use a rolling sphere with 1.4 Å radius with an integration step of 1 Å.
- The three output files generated by the program are currently named `provaly`, `provaly1` and `provaly2`, which correspond to the atomic coordinates; atomic radius, mass and surface accessibility²²; and correspondence i.e. atom name, residue name and residue name and number (Figure 5.1).

²² note that surface accessibility is determined for residue components i.e. whether it is a side chain or a main chain segment and not for individual atoms

HYDRODYNAMIC THEORY

ATOM	23	N	ASP	A	114	-5.242	15.453	-13.993	1.00	102.36	N
ATOM	24	CA	ASP	A	114	-5.926	15.898	-12.777	1.00	97.70	C
ATOM	25	C	ASP	A	114	-7.449	15.704	-12.726	1.00	90.41	C
ATOM	26	O	ASP	A	114	-8.233	16.466	-13.303	1.00	87.90	O
ATOM	27	CB	ASP	A	114	-5.518	17.341	-12.401	1.00	100.15	C
ATOM	28	CG	ASP	A	114	-4.073	17.435	-11.855	1.00	103.09	C
ATOM	29	OD1	ASP	A	114	-3.131	17.407	-12.676	1.00	105.26	O
ATOM	30	OD2	ASP	A	114	-3.876	17.549	-10.616	1.00	100.23	O
ATOM	31	N	LEU	A	115	-7.819	14.634	-12.027	1.00	85.90	N
ATOM	32	CA	LEU	A	115	-9.189	14.191	-11.781	1.00	84.95	C
ATOM	33	C	LEU	A	115	-8.982	13.076	-10.739	1.00	87.20	C
ATOM	34	O	LEU	A	115	-9.697	12.983	-9.731	1.00	82.64	O
27 0.00 provalyl											
-5.2420	15.4530	-13.9930	1.650	15	1	N	ASP	ASP_A_114			
-5.9260	15.8980	-12.7770	1.870	13	1	CA	ASP	ASP_A_114			
-7.4490	15.7040	-12.7260	1.760	12	1	C	ASP	ASP_A_114			
-8.2330	16.4660	-13.3030	1.400	16	1	O	ASP	ASP_A_114			
-5.5180	17.3410	-12.4010	1.800	14	1	CB	ASP	ASP_A_114			
-4.0730	17.4349	-11.8550	1.800	12	1	CG	ASP	ASP_A_114			
-3.1310	17.4070	-12.6760	1.800	16	1	OD1	ASP	ASP_A_114			
-3.8760	17.5490	-10.6160	1.800	16	1	OD2	ASP	ASP_A_114			
-7.8190	14.6340	-12.0270	1.650	15	1	N	LEU	LEU_A_115			
-9.1890	14.1910	-11.7810	1.870	13	10	CA	LEU	LEU_A_115			
-8.9820	13.0760	-10.7390	1.760	12	10	C	LEU	LEU_A_115			
-9.6970	12.9830	-9.7310	1.400	16	10	O	LEU	LEU_A_115			
-9.820	13.6230	-13.0610	1.800	14	1	CB	LEU	LEU_A_115			
-11.352	13.6530	-13.1370	1.800	13	1	CG	LEU	LEU_A_115			

Figure 5.1 Extract from a PDB file of fibrinogen, fragment D (first box) and the *BeaMS* formatted files in the three columns of the second box. The first column block contains only the x, y and z coordinates for each atom. The first value on the first line of this column block (which in this case is 27) indicates the number of lines to be read in by the program. The second value, 0.00 relates to the formatting of the correspondence file displayed in the third column block and was used in all cases considered in this study. The third value is the name of the rmc file, which is the file in the second column block. rmc stands for radius, mass and colour and corresponds to each of the three values in that column block i.e. the first and second value being the atomic radius and atomic mass while the third value indicates the whether the residue component is at the surface or buried. A value of 1 indicates a surface side chain while a value of 10 indicates a buried residue component. The value of 6 which is not shown in this column block indicates a surface peptide bond. The colour (or surface accessibility) information is the only new value to be added to the information given by the PDB file. The correspondence file, which is found in the third column contains the atom and residue names and residue number. Each line of each *BeaMS* file corresponds to a particular atom. However, when it is processed by *SOMO* each line corresponds to a bead. The mass information for each atom is ignored by *SOMO* whereas the mass information for each bead is processed by other *BeaMS* programs.

ASA can also be used to calculate the surface accessibility of bead models (option 3 - recheck), which requires a percentage value for the surface threshold. This is essentially a ratio of exposed surface area/total surface area for each bead. For example, if a value of 50% is used, then the rolling sphere has to cover at least half the area of the sphere for it to be tagged as being surface accessible. Therefore higher percentage values will result in a greater number of buried beads. The surface accessibility calculations for atomic models uses instead a fixed value of 10 \AA^2 , as originally proposed by Lee and Richards (1971). This feature is used to recompute the surface accessible beads and is commonly used for models containing many beads (> 500 residues).

By using only the surface accessible beads in the computation, the hydrodynamic parameters for the model can be calculated much more quickly. This should not be confused with a shell model just because it does not include 'buried' beads. In shell models the beads are much smaller and they are more closely packed along the surface. In fact, these medium-resolution models lie somewhat between shell and low resolution bead models (several residues per bead), and completely excluding the buried beads from the computations may lead to some errors. This was extensively tested, and a solution to this problem will be reported later.

5.2.2 SOMO

SOMO reads the three *ASA* output files (provaly, provaly1, provaly2) and an additional file containing the look up data table-provaly3 (chapter 4, Figure 4.15) file which mostly contains information about each residue and how it is to be processed by the *Trans* routine. The program is launched by calling *SOMO* with the provaly file name e.g. `somo T-lyso`.

The first option is selecting the algorithm choice:

- 0 - for *Trans*
- 1 - for *AtoB*

In version 1 (v1) of *SOMO*, if 0 is selected, four options are offered for processing each stage of the model. They are: (i) only remove overlaps; (ii) only pop beads; (iii) remove overlaps and pop beads; and (iv) do neither. The first stage of options (i) and (iii) will also prompt the user for whether outward translation of the surface side chain beads is required, while options

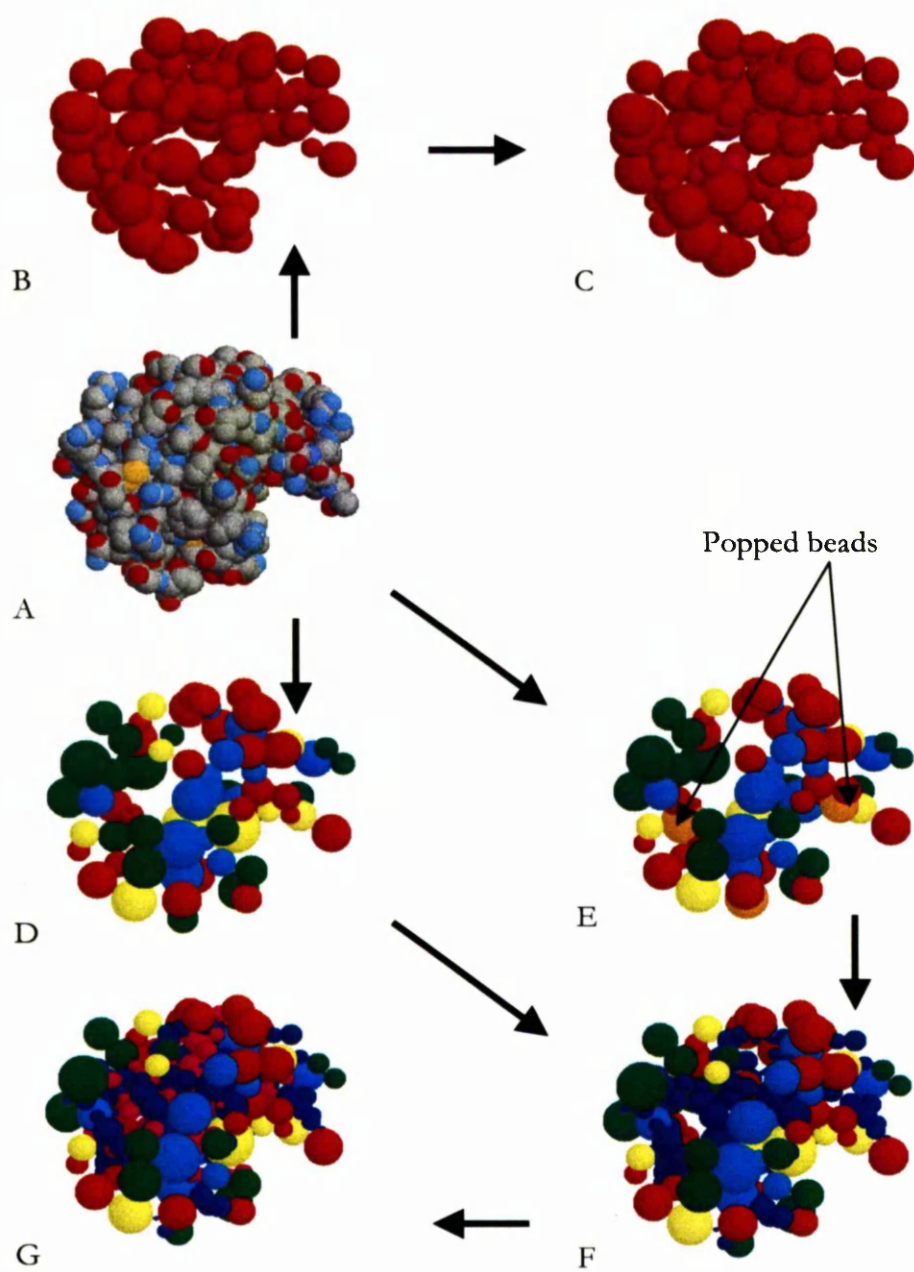


Figure 5.2 Bead processing stages in both *AtoB* (B and C) and *Trans* (D to G) algorithms, illustrated for lysozyme. In *AtoB*, using a grid spacing of 5 Å, surface beads are placed first (in red) with their overlaps removed and with outward translation, followed by the buried beads (in magenta). In *Trans* surface side chain beads are placed first (step E is the same as step D but with popping, as denoted by the orange beads), followed by surface peptide beads (in blue) (F) and finally buried beads (in magenta) (G). Colour codes for the side chain beads in *Trans* are: yellow, which indicates exposed acidic side chains; green, indicating exposed basic side chains; red, for exposed polar side chains and cyan, for exposed nonpolar or hydrophobic side chains.

(ii) and (iii) will prompt the user for the percentage overlap, above which beads are to be popped. Except for the case of CS, where 70% was used, the other models used a 20% overlap threshold i.e. if the overlapping volume between two beads is greater than 20% of the total volume of either bead, then the bead is popped. Popping, like overlap removal follows a specific order of events outlined by Figure 5.2 (see also previous chapter). The same four options are also offered if 1 (*AtoB*) is selected as is the additional request for providing a value for the spacing (in Å) of the cubic grid.

Four output files are generated by *SOMO*: output (an option is available for changing the default name), output.rmc, output.rmc1 and INPUT.DAT. Except for the last file (which is formatted for *HYDRO*), the formats are compatible for running with *RAZ*, *SUPC* and some of the other program modules within *BeaMS*.

5.2.3 RAZ

The bead models generated by *SOMO* were displayed using the molecular visualisation program *Rasmol* (Sayle and Milner-White, 1995). Because *Rasmol* does not read x, y, z, r files directly, two script files (.bms and .spt e.g. lyso.bms and lyso.spt) generated by the program *RAZ* are required (see enclosed CD). *RAZ* generates the scripts and then launches *Rasmol*.

5.2.4 SUPC

SUPC belongs to the *COEFF* group of programs, which includes *DLAG* and *GS*, all of which are used to calculate hydrodynamic parameters from bead models. The latter two programs

perform their computations using the Gauss-Siedel iterations (discussed by Spotorno *et al.*, 1997), whereas *SUPC* uses the full supermatrix inversion (see chapter 3). The hydrodynamic parameters for the models tested in this chapter were determined using *SUPC*.

A “readme” file accompanying the program describes in detail how to set up and run the program. In brief, at the beginning the program offers the choice of reading a single or multiple files. In the latter case, it will always generate a file containing the results of the output, as well as the average values of the hydrodynamic properties over all the models. The output file is optional for single models but quite necessary if the program is run on a PC under Windows, due to the lack of scrolling of the DOS window for output printed directly onto the screen. In the case of PDB-derived models, the coordinates from the input file are in Å while calculations by all the *COEFF* programs are in nm. Therefore a conversion factor of 0.1 is used to convert Å into nm. The hydrodynamic computation is performed either from the Cartesian origin or diffusion centre of the model, but only the latter was used in these tests. The stick boundary conditions were chosen over the slip, because they probably better reflect the interaction of the residue-sized beads with solvent. Volume and mass corrections were performed by summing the mass and volume of the beads as recommended by Carrasco and Garcia de la Torre (1999) for calculating the rotational diffusion coefficients and intrinsic viscosity of the molecules.

5.3 RESULTS

5.3.1 BEAD MODELS TESTED AT DIFFERENT WATER VOLUMES AND AT DIFFERENT RESOLUTIONS

A comparison of the hydrodynamic computations for bead models generated for three test proteins (not CS) and their literature experimental values are reported in Table 5.1. The first test which used the *Trans* routine compared the effect of using different values for the water molecule volume (Gerstein and Chothia, 1996) on the hydrodynamic properties of the models, popping those beads that overlapped greater than 20% (Table 5.1A). The intention of using the popping operator was to limit the extent of overlap removal in order to prevent the formation of very small beads. *SOMO* assumes a fixed number of water molecules associated with each residue *component*. For example, with *Trans* all main-chain segment beads have one water molecule associated with it, whereas a lysine side-chain bead will have 5 water molecules associated with it. Exact numbers used in this study can be found in the look-up data table (Figure 4.15). The values used were determined by Kuntz and Kauzmann (1974) by

NMR using frozen solvent to identify water protons in the vicinity of the macromolecule. In this study, no attempt was made to see the effect of varying the hydration level of individual residues. Instead, the *volume* assigned to the water molecules was changed to reflect the different volumes that have been measured for water (Gerstein and Chothia, 1996).

Three different water volumes of 20 \AA^3 , 24 \AA^3 and 30 \AA^3 were tested. The 24 \AA^3 volume corresponds to a weighted average between the mean electrostricted volume at the surface (24.5 \AA^3) and the mean for buried molecules (22.9 \AA^3) (Gerstein and Chothia, 1996) and is slightly lower than the 24.5 \AA^3 used by Perkins (1998). 20 \AA^3 and 30 \AA^3 are values situated on either side of the weighted electrostricted value. In fact the latter value of 30 \AA^3 corresponds to the average volume of bulk water, which is 29.7 \AA^3 ²³. To simulate the effect of no hydration, a volume of 0 \AA^3 was used. Data from these models were compared with models that did not use the popping routine (Table 5.1B). A third test using *AtoB* to generate models at 1 \AA and 5 \AA resolution with outward translation and at 5 \AA without translation was compared with models generated by *Trans* at an equivalent volume of water of hydration (24 \AA^3) (Table 5.1C). These three tests compare predicted with experimental parameters and how they change with different volumes of water of hydration and with different processing operators. In the final test, solution properties of models for all four proteins generated by *Trans* were run with and without: (i) the inclusion of buried beads, (ii) the recheck operator on *ASA* (iii) and with and without the inclusion in the volume correction (Garcia de la Torre and Carrasco, 1998) of the volume of the buried beads. The radius of gyration reported in both tables was computed from the crystal structure and was used as a check on the overall dimensions of the model.

Asynchronous overlap removal and outward translation (see previous chapter for details) was applied to all models generated by *Trans* and only to those not denoted by a T[§] in Table 5.1 with *AtoB*. The “popping” routine, allowing for the fusion of couples of beads if either bead overlaps by a volume greater than a specified threshold limit was set at 20% for all models running with the *Trans* algorithm except CS where 70% was used. 20% had been applied to CS, but this resulted in crashing the program as a result of excessive popping. The new value was chosen as a precaution against excessive bead shrinking. No attempt was made to investigate the hydrodynamic properties of models with overlapping beads of equal radii as this has been shown to give very poor values for intrinsic viscosity (Carrasco *et al.*, 1999b). A visual comparison of models (not to scale) generated by *Trans* and *AtoB* is given in Figure 5.3.

²³ The molar volume of water at 4°C is 29.9 \AA^3 which reduces to 29.5 \AA^3 at 50°C (Perkins, 1986).

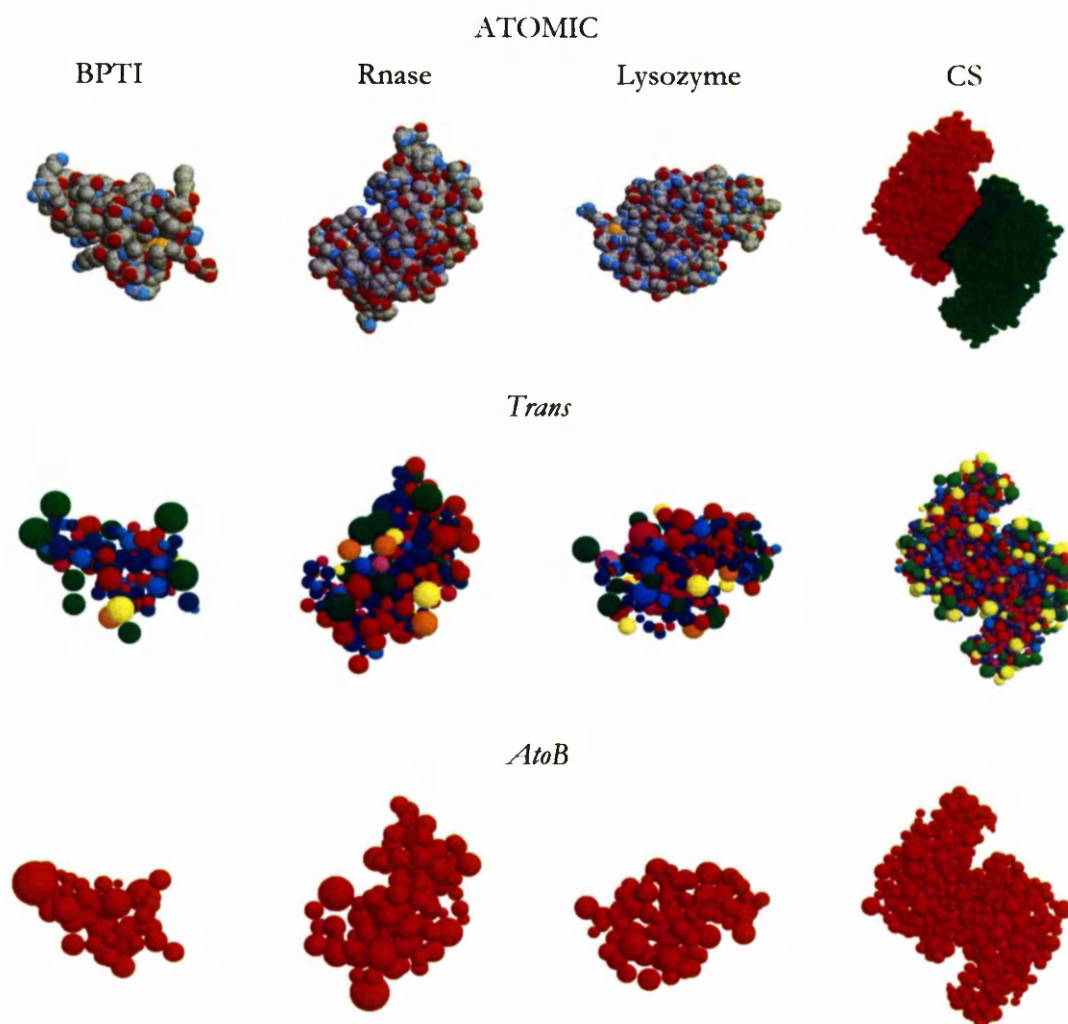


Figure 5.3 Atomic models for each of the four proteins: BPTI, RNase, lysozyme and CS are shown in the four columns of the first row. In the second row, each model was generated by *Trans* with outward translation and popping set at 20% for all the proteins except CS which was set at 70%. The last row represents models generated by *AtoB* with no outward translation, using a grid spacing of 5 Å. All the bead models used a water volume of 24 Å.

Taking the *Trans* algorithm first (Table 5.1A and B); the application of the popping routine resulted in greater numbers of beads being popped as the solvent volume was increased. This was to be expected considering that beads were being enlarged by the increase in solvent volume and hence were more likely to have larger overlaps. The outcome of too much popping was a lower resolution model. In these tests, the maximum number of popped beads was 31 (from the RNase model), which constituted less than 15% of the total number of beads in that model. In any case, the popping had a relatively minor effect on the most shape sensitive parameters such as rotational correlation time, τ_c^h , rotational diffusion coefficient,

D_t , and intrinsic viscosity, $[\eta]$. This came as a relative surprise, considering that popping changes the model not only by reducing the number of beads in the model but also by altering the order in which bead overlaps are removed. This is because overlaps are removed according to an overlap hierarchy with the largest overlaps being removed first. So a popping event is most likely to alter this hierarchy. Evidently, the way in which the popping is performed does not alter too much the overall “geometry” of the model, and it can be safely applied if limited to a few beads.

For all three proteins, the models perform quite well and the influence of the water volume is relatively minor, with the rotational dynamics being the most effected. The models for each of the different water volumes are in fact visibly indistinguishable as shown by Figure 5.4. The Stokes radii R_s of the models of the two bigger proteins tested (RNase and lysozyme) were biased towards the lower end of the range of experimentally derived values. As this parameter is affected mostly by the overall dimension (elongation) of the molecule, it may indicate that the models are smaller than they should be. The τ_c^h for all the models at all levels of hydration fell within experimentally determined values. As for the intrinsic viscosity, that of RNase lies outside the range of experimentally derived parameters for the hydrated models. Its value was only improved slightly by the absence of popping. As for lysozyme, the range of the experimentally-derived value is so large that it is difficult to assess the performance of the models, even if their values fell within it. In all three cases (with and without popping) higher R_g values were reported. This was unsurprising considering that those models were generated with outward translation of the surface beads during overlap removal thus altering the mass distribution in the model. The R_g values reported in Table 5.1C (for models generated by *AtoB*) without translation gave almost exact values to those predicted from the crystal structure. However, the remaining predicted parameters were found to be much poorer than those generated with the *Trans* method. For a start, the *AtoB* models were visibly less compact (Figure 5.3), particularly when outward translation was applied. With translation, the predicted parameters make a general improvement, only compromising on R_g . It would appear that the effect of translation has a much greater impact on the hydrodynamic parameters than the change in resolution from 1 Å to 5 Å (Figure 5.5). This is unsurprising considering that a resolution of 1 Å is very high and is unlikely to differ too much from a 5 Å resolution model. To re-emphasise, the reason for using 1 Å resolution models was to ensure that each residue was only represented by a single bead.

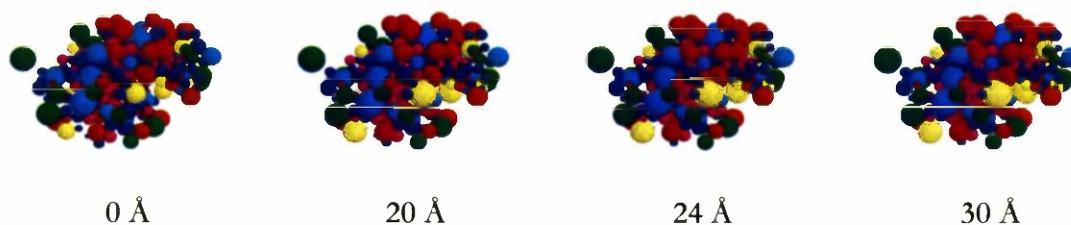


Figure 5.4 Models for lysozyme generated by *Trans* using no hydration (this was simulated using a water volume of 0 Å), and then water volumes of 20 Å, 24 Å and 30 Å. Outward translation was applied to all the models, whereas popping was not. Each of these models looks almost visibly indistinguishable from the others.



Figure 5.5 *AtoB* generated models for BPTI, at 1 Å resolution without outward translation (1) and at 5 Å resolution without (2) and with outward translation (3).

Table 5.1 A

BPTI (454 atoms)	Experimental value	W: 0 Å ³ B#: 109	W: 20 Å ³ B#: 109	W: 24 Å ³ B#: 107	W: 30 Å ³ B#: 107
R_e, nm	1.39 ^a -1.49 ^b -1.52 ^c	1.41	1.45	1.46	1.47
D_r, s⁻¹ × 10⁷	n.a.	4.14	3.82	3.74	3.68
τ_c^h, ns	4.1 ^d -4.8 ^e	4.10	4.45	4.53	4.61
[η], cm³/g	n.a.	3.45	3.76	3.83	3.90
R_g, nm	1.12	1.15	1.15	1.15	1.16
RNase (956 atoms)	Experimental value	W: 0 Å ³ B#: 243	W: 20 Å ³ B#: 232	W: 24 Å ³ B#: 232	W: 30 Å ³ B#: 213
R_e, nm	1.77 ^f -1.91 ^g	1.77	1.79	1.80	1.80
D_r, s⁻¹ × 10⁷	n.a.	2.15	2.01	2.00	1.99
τ_c^h, ns	8.34 ^h -9.2 ⁱ	7.89	8.43	8.48	8.51
[η], cm³/g	3.16 ^j -3.30 ^k	3.25	3.50	3.52	3.55
R_g, nm	1.45	1.46	1.46	1.46	1.46
Lysozyme (1001 atoms)	Experimental value	W: 0 Å ³ B#: 240	W: 20 Å ³ B#: 236	W: 24 Å ³ B#: 234	W: 30 Å ³ B#: 225
R_e, nm	1.82 ^l -1.85 ^m -1.90 ⁿ	1.78	1.80	1.82	1.81
D_r, s⁻¹ × 10⁷	1.7 ^o -2.2 ^p	2.13	2.02	1.95	1.97
τ_c^h, ns	8.3 ^q -8.4 ^r -8.9 ^s	7.97	8.39	8.68	8.63
[η], cm³/g	3.0 ^t -3.6 ^u	3.17	3.32	3.46	3.43
R_g, nm	1.42	1.44	1.44	1.44	1.45

Table 5.1 B

BPTI (454 atoms)	Experimental value	W: 0 Å ³ B#: 109	W: 20 Å ³ B#: 109	W: 24 Å ³ B#: 109	W: 30 Å ³ B#: 109
R_e, nm	1.39-1.49-1.52	1.41	1.45	1.46	1.47
D_r, s⁻¹ × 10⁷	n.a.	4.14	3.80	3.75	3.68
τ_c^h, ns	4.1-4.8	4.10	4.47	4.53	4.60
[η], cm³/g	n.a.	3.45	3.78	3.83	3.91
R_g, nm	1.12	1.15	1.15	1.16	1.16
RNase (956 atoms)	Experimental value	W: 0 Å ³ B#: 244	W: 20 Å ³ B#: 244	W: 24 Å ³ B#: 244	W: 30 Å ³ B#: 244
R_e, nm	1.77-1.91	1.77	1.79	1.80	1.80
D_r, s⁻¹ × 10⁷	n.a.	2.15	2.03	2.02	1.99
τ_c^h, ns	8.34-9.2	7.89	8.33	8.39	8.50
[η], cm³/g	3.16-3.30	3.25	3.44	3.46	3.51
R_g, nm	1.45	1.46	1.46	1.46	1.46
Lysozyme (1001 atoms)	Experimental value	W: 0 Å ³ B#: 245	W: 20 Å ³ B#: 245	W: 24 Å ³ B#: 245	W: 30 Å ³ B#: 245
R_e, nm	1.82-1.85-1.90	1.78	1.80	1.80	1.81
D_r, s⁻¹ × 10⁷	1.7-2.2	2.13	2.06	2.02	2.00
τ_c^h, ns	8.3-8.4-8.9	7.97	8.34	8.41	8.50
[η], cm³/g	3.0-3.6	3.17	3.32	3.35	3.39
R_g, nm	1.42	1.44	1.44	1.44	1.44

Table 5.1 C

BPTI 6,492 Da	Experimental value	R: 1 Å³ B#: 58	R: 5 Å³ B#: 48	R: 5 Å³ B#: 48 T^{\$}
R_e, nm	1.39-1.49-1.52	1.33	1.34	1.25
D_r, s⁻¹ × 10⁷	n.a.	4.85	4.62	5.62
τ_c^h, ns	4.1-4.8	4.11	4.29	3.50
[η], cm³/g	n.a.	2.97	3.12	2.56
R_g, nm	1.12	1.19	1.2	1.11
RNase (13,657 Da)	Experimental value	R: 1 Å³ B#: 124	R: 5 Å³ B#: 101	R: 5 Å³ B#: 101 T^{\$}
R_e, nm	1.77-1.91	1.69	1.7	1.67
D_r, s⁻¹ × 10⁷	n.a.	2.42	2.39	2.48
τ_c^h, ns	8.34-9.2	7.61	7.75	7.36
[η], cm³/g	3.16-3.30	2.84	2.85	2.76
R_g, nm	1.45	1.49	1.5	1.45
Lysozyme (14,279 Da)	Experimental value	R: 1 Å³ B#: 129	R: 5 Å³ B#: 111	R: 5 Å³ B#: 111 T^{\$}
R_e, nm	1.82-1.85-1.90	1.73	1.80	1.65
D_r, s⁻¹ × 10⁷	1.7-2.2	2.25	2.28	2.54
τ_c^h, ns	8.3-8.4-8.9	8.45	8.32	7.38
[η], cm³/g	3.0-3.6	2.96	2.92	2.63
R_g, nm	1.42	1.49	1.44	1.41

^aDynamic light scattering (Wills and Georgalis, 1981); ^bDynamic light scattering (Gallagher and Woodward, 1989); ^cPulsed-field gradient NMR (Ilyina *et al.*, 1997); ^dFrequency-domain intrinsic fluorescence polarization anisotropy (Lakowicz *et al.*, 1987); ^eNMR relaxation (Otting *et al.*, 1991); ^fSedimentation velocity (Rothen, 1940) calculated from $Dt = 1.36 \times 10^{-6} \text{ cm}^2/\text{s}$; ^gDynamic light scattering (Rimai *et al.*, 1970); ^hElectric birefringence (Krause and O'Konski, 1963); ⁱDielectric

relaxation (Keefe and Grant, 1974); ^j Deduced from $R_{\eta} = 1.9$ nm from table I in Dubin *et al.* (1993); ^k The reported value is an average taken over various different conditions (such as pH) (Buzzell and Tanford, 1956); ^l Interferometry (Luzzati and Champagne, 1957); ^m Dynamic light scattering (Grigsby *et al.*, 2000; Nicoli and Benedek, 1976); ⁿ From sedimentation and diffusion coefficients measured by sedimentation velocity (Dubin *et al.*, 1971); ^o Fabry-Perot interferometry (Dubin *et al.*, 1971); ^p Depolarised dynamic light scattering (Chirico *et al.*, 1999); ^q NMR relaxation (Buck *et al.*, 1995); ^r Steady state extrinsic fluorescence depolarization (Steiner and McAllister, 1957); ^s NMR relaxation (Kakalis and Baianu, 1988); ^t (Sophianopolous *et al.*, 1962); ^u From Table I in Dubin *et al.* (1993).

Table 5.1 Comparison between experimental and calculated values for the three test proteins: BPTI, RNase and lysozyme. In tables A and B, the comparison is as a function of water volume using *Trans* generated models with outward translation and popping set at 20% (only in table A) and no popping (only in table B). Table C compares models generated by *AtoB* at 1 Å and 5 Å resolutions with outward translation and at 5 Å resolution without outward translation. The number of beads in the model is indicated by B#; W indicates the water volumes used (0, 20 Å, 24 Å and 30 Å); R the resolution of the model and T\$ indicates no outward translation. The radii of gyration were computed directly from the crystal structures. The bracketed terms in table A and B are the number of atoms used in the model, whereas in table C, they refer to the mass of the molecule as calculated from its amino-acid sequence.

5.3.2 EFFECT OF BURIED BEADS AND VOLUME CORRECTIONS ON HYDRODYNAMIC PARAMETERS

The 'rmc' file generated by *SOMO* is particularly important for identifying beads to be used in the computation. Because hydrodynamic theory emphasises the significance of those beads interacting directly with the bulk solvent as being responsible for the solution parameters of the model, the elimination of the buried beads should not make a big difference to the hydrodynamic calculations. The *ASA* coding which is used to distinguish between buried and exposed surface side chain and peptide beads can be used by *SUPC* for including in the computations only the surface accessible beads i.e. by excluding (from the hydrodynamic computation) beads colour coded "6". The only other colour coding besides "6" which will automatically result in a bead being excluded from hydrodynamic computation is "0". This value is assigned by *SOMO* to beads whose size is less than 0.001 Å. Such small beads can arise as a result of overlap removal. Because of their very small size, they were viewed as unimportant in the hydrodynamic computations and therefore worthy of exclusion from the

supermatrix computation. *SUPC* upon reading the *SOMO* output files will alert the user of the presence of these small beads while excluding them only from the computation. The computations will still however retain the mass information for these small beads so that there is no loss of mass in the complete model. For normal use as in the above tests, all beads were included in the computation. However, the effect of excluding them from the computation was found to have minimal effect particularly for those parameters not so sensitive to shape, as presented in Table 5.2 (the initial model column).

A large proportion of beads from each model although labelled as buried, also appeared to be exposed to solvent sized molecules. For example, the BPTI model resulting from the *Trans* routine comprises 109 beads of which 94 belong to exposed residues as determined by *ASA*. But the packing in this model is such that even those few beads that are labelled as buried also appear to be exposed. The reason for this is simply that the beads were labelled according to the solvent accessibility of the residue components from the atomic model and not that of the current model. The transformation from atomic to a bead model with no overlaps will inevitably result in a more loosely packed model, exposing residue components previously labelled as buried. The re-classification of each bead according to solvent accessibility is done by the recheck routine in *ASA*.

In the example of BPTI, the recheck routine using the parameters outlined in the methods section identified 99 exposed beads i.e. 5 beads more than previously categorised. In Table 5.2, the effects of using all the beads from the initial model, or only the exposed ones has minimal difference on the hydrodynamic parameters. This difference is even less for those models that have been rechecked. In any case, the differences in predicted parameter values are insignificant when contrasted with the wide ranging values for the experimentally measured parameters.

A further test on the same models (generated by *Trans*) was used to investigate the effect of adding the volume of the buried beads (while excluding them from the hydrodynamic computations) to the volume corrections suggested by Garcia de la Torre (1998) for the rotational diffusion and intrinsic viscosity. All the models described so far have included the volume of the buried beads in the volume correction, but removing it, as shown in the final column of Table 5.2, did result in a difference in values. In the case of the rotational diffusion, not including the volume correction produced worse values, while for the intrinsic viscosity it is harder to judge. Take RNase for example; by *not* using the full volume correction, the predicated intrinsic viscosity value was brought closer to the experimental range of values.

		Initial model		Rechecked	
		All beads	Exposed	Exposed	Exposed
BPTI (454 atoms)	Exp.	B#: 109 (u: 109, c)	B#: 109 (u: 94, c)	B#: 109 (u: 99, c)	B#: 109 (u: 99, nc)
R_e, nm	1.39-1.52	1.46	1.46	1.46	1.46
D_r, s⁻¹ × 10⁷	n.a.	3.75	3.75	3.75	3.77
τ_c^h, ns	4.1-4.8	4.53	4.52	4.53	4.50
[η], cm³/g	n.a.	3.83	3.81	3.82	3.80
R_g, nm	1.12	1.15	1.16	1.15	1.15
RNase (956 atoms)	Exp.	B#: 244 (u: 244, c)	B#: 244 (u: 180, c)	B#: 244 (u: 205, c)	B#: 244 (u: 205, nc)
R_e, nm	1.77-1.91	1.80	1.79	1.80	1.80
D_r, s⁻¹ × 10⁷	n.a.	2.02	2.02	2.02	2.05
τ_c^h, ns	8.34-9.2	8.39	8.37	8.39	8.27
[η], cm³/g	3.16-3.30	3.46	3.45	3.46	3.41
R_g, nm	1.45	1.46	1.46	1.46	1.46
Lysozyme (1001 atoms)	Exp.	B#: 245 (u: 245, c)	B#: 245 (u: 166, c)	B#: 245 (u: 190, c)	B#: 245 (u: 190, nc)
R_e, nm	1.82-1.9	1.80	1.80	1.80	1.80
D_r, s⁻¹ × 10⁷	1.7-2.2	2.02	2.02	2.02	2.06
τ_c^h, ns	8.3-8.9	8.41	8.40	8.41	8.23
[η], cm³/g	3.0-3.6	3.35	3.32	3.34	3.27
R_g, nm	1.42	1.44	1.44	1.44	1.44

Table 5.2 Predicted parameters from models generated by *Trans*, using all the beads; just the exposed beads from the initial model and then from models that have been rechecked by *ASA*, with (c) and without (nc) summing the volume of the buried beads to the volume correction. All four different types of models were compared with experimental data (**Exp.**), references for which can be found in the previous table.

The hydrodynamic behaviour of a large protein was also investigated by applying both *Trans* and *AtoB*. Citrate synthase (CS), which is 97.7 kDa in size, was chosen. Values for sedimentation and diffusion coefficients and for intrinsic viscosity have been compiled by Zipper & Durchschlag (2000). In Table 5.3, the computed values are given for a model generated with a water volume of 24 Å³ and a popping threshold of 70% in the case of *Trans* generated models (with recheck and with and without the addition of the volume of the buried beads to the volume correction) or a 5 Å resolution model with *AtoB*. The parameter values were then compared with experimental ones. A very good agreement is found with the *Trans* generated models, be it with or without inclusion of buried beads. Also, unlike RNase, the intrinsic viscosity appears to get *worse* without the inclusion of the volume of the buried beads in the volume correction. The *AtoB* models while giving precise values for R_g , fare less well with the other parameters.

It should be recalled that Carrasco and Garcia de la Torre (1999) have pointed out that the volume correction should not be applied to bead models of the “filled” type. The models generated by *SOMO* are somewhere in between filled shell and “whole body” bead models, and therefore the full application of the volume correction for the intrinsic viscosity will have to be evaluated separately. The significance of this is that for large proteins such as CS (Table 5.3) which generates 1682 beads, hydrodynamic computations performed only on those beads rechecked as being surface accessible (1093) saved a considerable amount of computation time.

97,690 Da (3445 atoms)	Exp.	T* B#:1682			A* B#:749
		(u:1682, c)	(u:1093, c)	(u:1093,nc)	(u: 749, c)
s, (S)	6.2	6.15	6.15	6.15	6.49
D _t , cm ² s ⁻¹ × 10 ⁻⁷	5.8	5.87	5.87	5.87	6.18
[η], cm ³ /g	3.95	3.89	3.90	3.75	3.36
R _g , nm (Crystl.)	2.81	2.83	2.83	2.83	2.81
R _g , nm (SAS)	2.91	2.83	2.83	2.83	2.81

Table 5.3 Comparing Experimental (Exp.) data for CS (obtained from Zipper and Durschlag (2000) with models generated by *Trans* (T^u) (70% popping) and *AtoB* (A^u) (5 Å resolution). u indicates the number of beads used for computing the hydrodynamic parameters while c and nc corresponds to corrected and uncorrected volume calculations.

5.4 DISCUSSION

The results show quite clearly the viability of the *Trans* algorithm for predicting hydrodynamic parameters. In almost all cases, the predicted parameters were either within the range of experimental measurements or just outside it. The *AtoB* algorithm does not fare as well, probably because it generates a slightly lower resolution model which limits then the extent to which it can model the surface irregularities of the molecule. Visualising the models generated by *AtoB* shows that it does not have the close packing arrangement found in *Trans* generated models (this is seen more clearly from the computer model than the images in Figure 5.3. Clearly packing is bound to be less efficient for models with no overlaps particularly as the resolution increases, but whether this is the only reason for the poor packing is not known. An alternative method for removing overlaps whose implementation has not yet been tested is that of synchronous overlap removal. This method identifies the two overlapping beads with the smallest overlap and then proceeds to reduce the radii of all overlapping beads by this amount. The testing of this procedure is planned in the near future and will provide an alternative if not improved procedure for overlap removal.

Although the current version of *Trans* is an improvement to previous versions and fares better than the *AtoB* algorithm, it is much more restrictive for lowering the resolution of the model. While *AtoB* can only attain a maximum resolution of one residue per bead, it can change its grid spacing to generate much lower resolution models. While this may be appealing for modelling much larger structures (which have yet to be tested), care should be taken that this does not result in any compromise in the accuracy of the predictions. *Trans* can lower its resolution by lowering the popping threshold so that more beads are popped. Initial studies (data not shown) on the three small proteins indicate that popping threshold values less than 15% were found to cause a cascade of popping events²⁴. Higher threshold values would be required to prevent this also from happening with CS. It should also be pointed out that the popping routine was developed with the intention of fusing beads with large overlaps. Although in this study small overlap thresholds have been used, it is likely that its continued use will be for fusing beads with much larger overlaps as in the CS model.

While accurate modelling of the hydration shell is important, its impact on non-shape sensitive parameters was found to be minimal as seen in Table 5.1. Even the shape sensitive parameters were not too affected by the volume of the associated water molecules. However, the effects of directly varying the hydration level by changing the number of the water molecules associated with each residue have not been tested, so further tests will be needed. Part of the problem with gauging the accuracy of these predictions is the difficulty in obtaining accurate experimental measurements. A study carried out by Perkins (1986) showed that the partial specific volumes of lysozyme and ribonuclease A as determined from its sequence composition compared poorly with densimetry measurements especially when related to other test proteins used in the same study. The reason was attributed to both proteins being more hydrated than the other test proteins. Evidence for such a conclusion was shown by Bull (1981) cited in Perkins (1986) and by Keefe and Grant (1974) in the case of ribonuclease A which was determined to have a 0.8 nm thick hydration shell.

Although the results from the tests are promising, only a few models were tested and except for CS they were all relatively small and similar in size (between 6 – 15 kDa). Further tests are being planned to test molecules that are elongated, large, composed of multiple subunits and flexible. Except for elongated molecules, the outcome should not be too different from the precision obtained for the test systems. The drawback for elongated molecules is that outward translation occurs from the centre of gravity as opposed to the plane of the surface.

²⁴ The cascade effect mostly occurs when two large beads are being popped which then results in those beads occupying a much larger volume resulting in the popping of even more beads. Because larger molecules are likely to undergo more popping events, there is a greater likelihood of two popped beads (hence relatively large beads) being popped.

A solution to this is proposed in chapter 4 but it has not been implemented in the current version of *SOMO*.

One possible improvement to *AtoB* would be to hydrate all the beads in the model as happens in *Trans*. The premise of hydrating all the beads was simply that all the residues regardless of location had a likelihood of being hydrated, so that location or surface accessibility only determines the order in which bead overlaps are removed. The results of this study appear to suggest that the overall dimensions may be a little too compact. Zamyatnin (1972) estimated the protein volume to be bigger than the sum of the volumes of those elements composing it. This suggestion was explained by the fact that solvent interacting with the exposed residues of an unfolded protein would be electrostricted. But then as the protein folds, interactions with other residues will decrease the charge attraction with water causing it to expand slightly thus increasing the effective volume of those residues and hence the overall volume. This view was supported by data presented by other workers such as Linderstrom-Lang (1962) which is cited in Zamyatnin (1972).

The current trend in bead model generation is towards much higher resolution models. Another feature not yet completed in *SOMO* is the generation of bead models using a similar procedure as *AtoB* but placing a grid over the actual atoms as opposed to residues (converted from the atomic data). This would help raise the resolution so that each residue could be represented by more than 1 bead and the beads would still be positioned according the mass distribution in the model. It would therefore be an improvement over the current implementation of *HYDROPRO* (which essentially begins by creating a shell model and then filling it with non-overlapping beads of equivalent size i.e. does not account for the mass distribution within the model), particularly for predicting SAS curves such as $I(Q)$ versus Q and $p(r)$ versus r .

This study forms an initial test of some of the features currently implemented in *SOMO*. More work is planned particularly for comparing parameters determined in SAS experiments. While the *Trans* algorithm in *SOMO* was developed specifically for hydrodynamic modelling, its suitability for SAS data is likely to be compromised by its positioning of the beads away from the c.o.g. for a cluster of atoms found at or close to the surface of the model. Whereas *AtoB* places beads more so at the centre of mass of a collection of atoms which should have less of a detrimental effect on the overall R_g and $p(r)$ of a molecule as observed in this present study. To model mass distribution in the *Trans* algorithm more accurately while still placing beads at positions best representing the location of residue components requires that the mass of each bead be independent of the collection of atoms that it represents. In this way the

mass can be more freely distributed among other beads thereby helping to improve this overall distribution.

CHAPTER 6

PRODUCTION, PURIFICATION AND DERIVATISATION OF PNEUMOLYSIN

6.1 INTRODUCTION

Pneumolysin is a membrane damaging, thiol-activated toxin which is 53 kDa in size and is expressed by the gram positive bacterium, *Streptococcus pneumoniae*. This bacterium is one of the most important causes of community acquired pneumonia (CAP) and is also responsible for otitis media, sepsis and meningitis. Experimental studies have shown the route of infection to begin with colonisation of the nasopharynx and then spreading either upwards through the Eustachian tube and into the middle ear to cause otitis-media or downwards into the alveoli of the lower respiratory tract to cause lobar pneumonia (Cundell *et al.*, 1995).

The toxin is expressed at low yields from *S. pneumoniae*, which is one of the reasons for inserting the pneumolysin gene into *Escherichia coli* (Mitchell *et al.*, 1989), which produces toxin at higher yields under optimal growth conditions. Comparison between pneumolysin obtained from *S. pneumoniae* and *E. coli* showed no obvious differences in terms of size, specific activity and inhibition by cholesterol (Mitchell *et al.*, 1989) and was therefore considered to be a viable method for producing toxin for further experimental studies. The purification strategy involved passing cell lysate through a hydrophobic interaction chromatography (HIC) column, where fractions containing pneumolysin were re-purified using the same column. This procedure was later modified by an orthogonal purification procedure i.e. replacing the second HIC step with an anion exchange chromatography (AEC) step (Gilbert *et al.*, 1998). Comparisons between pneumolysin obtained from *Streptococcus pneumoniae* and *E. coli* showed no obvious differences in terms of size, specific activity and inhibition by cholesterol (Mitchell *et al.*, 1989).

Pneumolysin is predominantly monomeric in solution (Morgan *et al.*, 1993) but is known to oligomerise in the presence of cholesterol both within the membrane or in solution to form ring and arc shaped structures (Morgan *et al.*, 1994) and helices (Gilbert *et al.*, 1999; Morgan *et al.*, 1995). It has also been recently shown that oligomerisation can occur spontaneously in deuterated phosphate buffer saline (PBS), in the absence of cholesterol (Gilbert *et al.*, 1998). The formation of oligomers in deuterated PBS was substantially reduced by derivatising the toxin with dithio(bis)nitrobenzoate (DTNB) (Gilbert *et al.*, 1998). The reaction (based on Wang *et al.* (1996)) involves the attachment of a benzyl moiety (Φ) from DTNB (Φ -S-S- Φ) via a disulphide bridge (S-S) to the thiol group on the single cysteine (SH) residue found close to the C-terminal region of the toxin (PLY) as shown by the mechanism below:

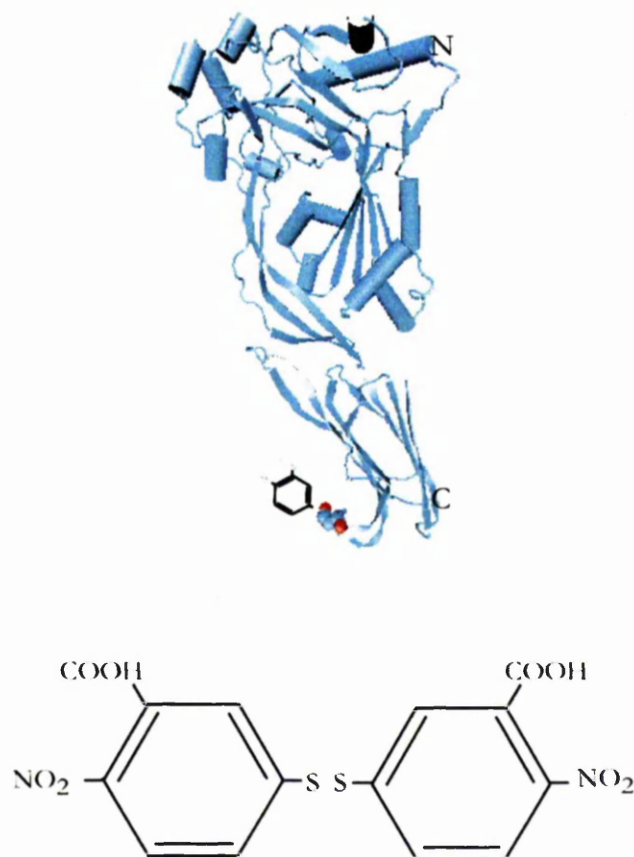
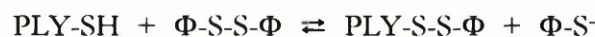


Figure 6.1 Homology model of pneumolysin (Rossjohn *et al.*, 1998) visualised by VMD (Humphrey *et al.*, 1996) showing the single cysteine residue as spheres near the C-terminus. Reduction of DTNB (pictured below the model) between the sulphur atoms causes it to split into two TNB molecules, one of which then binds to the cysteine residue.

Free TNB ($\Phi\text{-S}^-$) absorbs strongly at 412 nm, whereas the protein conjugated with TNB (PLY-S-S- Φ) absorbs at 337 nm (Wang *et al.*, 1996). AUC studies by Gilbert *et al.* (1998) on the derivatised sample indicated that a minority of toxin molecules present in solution had not been derivatised. This was supported by haemolytic tests, showing that the derivatised sample was only capable of inhibiting 90% of sheep red cell lysis whereas derivatised θ -toxin (from *Clostridium perfringens* type A), which also has a single cysteine residue near the C-terminus showed 100% inhibition of haemolysis²⁵ (Iwamoto *et al.*, 1987). Although the majority of

²⁵ Inhibition was caused by the inability of the derivatised toxin to bind to the erythrocyte membrane. The binding constant was approximately 1% of native toxin.

toxin was monomeric, AUC measurements using schlieren optics showed that a small proportion was dimeric. The cause for the latter was attributed to underivatised toxin as opposed to the TNB conjugate binding to a cysteine residue on underivatised pneumolysin. Small angle neutron scattering has been used extensively for structural studies of macromolecules, usually placed in a deuterated (D_2O) buffer. The ability to prevent the majority of toxin oligomerising in D_2O has been of significant benefit for studying the monomeric form of the toxin using SANS experiments.

The following study outlines a purification procedure using the HIC and AEC steps used by Gilbert *et al.* (1998), but replacing the Phenyl-sepharose and Mono-Q matrix (Pharmacia, UK) with POROS PE and HQ (Applied Biosystems, UK) respectively. Although the purification procedure used by Gilbert *et al.* (1998) was successful, it was of interest to see if the purification rate and yield could be further improved with this new media. POROS chromatographic media patented by PerSeptive Biosystems in 1989 (now known as Applied Biosystems) had paved the way for faster purification rates for biopolymer mixtures than most conventional media (Dughan, 1994). The results in this chapter show that purification of a 20 ml crude extract by HIC can take less than two hours and that the passage of sample through the matrix took less than six minutes. In total, three steps were used to purify the toxin prior to derivatisation: HIC, AEC and SEC (size exclusion chromatography), with only the latter step requiring the use of an FPLC (fast protein liquid chromatography) system from Pharmacia.

Sedimentation equilibrium analysis using the AUC revealed that the final sample was homogenous. The toxin was derivatised using the same procedure as Gilbert *et al.* (1998) but ensuring that the toxin concentration was very dilute (less than $1 \mu M$). Analysis by absorbance spectroscopy and mass spectrometry revealed that this improved both the quantity and proportion of derivatised toxin. This chapter will provide a brief introduction to perfusion chromatography and sedimentation velocity, followed by the experimental details for expressing, purifying and derivatising pneumolysin.

6.1.1 PERFUSION CHROMATOGRAPHY

Prior to perfusion chromatography, FPLC (pioneered by Pharmacia Biotech) and high pressure liquid chromatography systems (HPLC) were recognised as the standard methods for purifying biomolecules. FPLC was more popular for protein purification than HPLC because larger quantities of protein could be loaded onto the chromatography column (from 0.2 mg in HPLC to 20 mg in FPLC) and the column was not subject to such high pressures (from 55

MPa in HPLC to between 1- 2 MPa in FPLC) (Sheehan, 2000). The main difference between perfusion chromatography and FPLC is in the physical structure of the matrix (Figure 6.2), although there has been some dispute over this (Dughan, 1994). Both chromatographic techniques possess very small beads which are used to pack the column matrix. Perfusion chromatography uses POROS media that comprises of many small pitted microspheres made of poly(styrene-divinylbenzene) that are bonded together to form bead like structures. Depending upon the type of matrix, beads can be either 10, 20 or 50 μm in size. The beads in this particular study were 20 μm .

Ligand coated channels (600 to 800 nm in diameter) running through the bead allow molecules flowing through it to be either unhindered, retarded or stopped according to their affinity for the ligand. These channels not only increase the area for binding biomolecules but because they are narrow, the proximity between molecule and ligand²⁶ is much smaller. This is a critical factor in the POROS design because the binding of molecule to ligand is dependent on diffusion. By reducing the effective distance between molecule and ligand means that solution can be passed through the column at a faster rate without compromising too much on resolution. Although the purification rates can be faster, the pressure inside the system can also be higher than those typically encountered in FPLC systems, which may be a problem for purifying proteins, particularly if they have a tendency to aggregate.

²⁶ The ligands are phenyl-ether (PE) for the HIC and high density quaternary ethyl (HQ) for AEC.

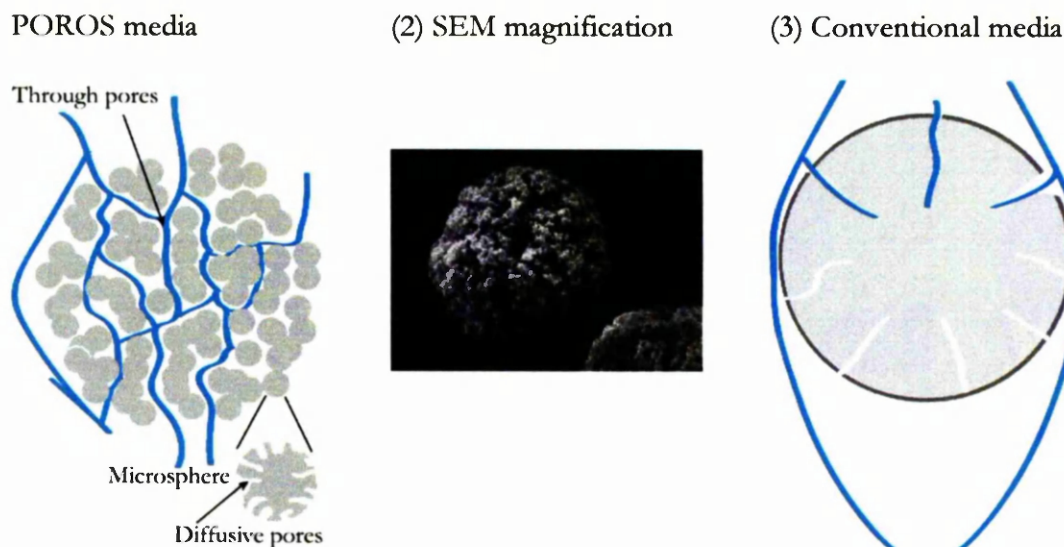


Figure 6.2 POROS media is manufactured by adhering microspheres into particle clusters (1). Solutions pass between spaces found in the cluster giving molecules an increased likelihood of reaching the surface of the microspheres by diffusion. The size of these spaces, also known as through-pores, ranges between 6000-8000 Å. Each microsphere is pitted with crevices (diffusive pores of 800-1500 Å), which increases the surface area even further for molecules to attach. (2) A scanning electron micrograph (SEM) of a single POROS bead at 5000 X magnification (taken from <http://www.appliedbiosystems.com/perseptive>). (3) Long internal crevices found in conventional media can result in molecules being trapped in stagnant mobile phase pools which can result in peak broadening and loss of capacity if flow rates are increased.

6.2 MATERIALS AND METHODS

Toxin production and purification were based on protocols initially developed by Mitchell (1989) and later modified by Gilbert (1998). All chemicals were obtained from Sigma-Aldrich company Ltd, except where stated. All chromatography buffers were steam sterilised and filtered through an AP20 (0.8 µm - 8 µm) glass fibre prefilter obtained from Millipore. Preparation of materials if not stated in this section can be found in Sambrook (1989). All procedures involving the handling of micro-organisms and toxins were conducted according to standard guidelines for good laboratory practice.

6.2.1 BACTERIAL STOCK PREPARATION

Escherichia coli strain M15, possessing the wild type pneumolysin gene cloned from the pneumococcal genome (Walker *et al.*, 1987) was provided by Professor T. J. Mitchell, University of Glasgow. The gene was inserted into an IPTG-inducible over expression vector (pkk) which carries resistance to both ampicillin and kanamycin (Amann and Brosius, 1985). Bacteria were grown on nutrient agar base (Oxoid, UK) containing ampicillin at a concentration of 100 µg/ml and kanamycin at a concentration of 20 µg/ml. Stock cultures were made using individual colonies selected from the agar base which were used to seed a 100 ml volume of Luria broth (LB Miller's broth) containing 100 µg/ml of ampicillin and 20 µg/ml of kanamycin. The cultures were grown overnight in a shaking incubator set at 220 rpm at 37°C. 1 ml aliquots of the culture containing 20% sterile glycerol were stored at -70°C until required.

6.2.2 PRODUCTION OF CRUDE TOXIN

1 ml of *E. coli* (M15pkkwt) stock culture was added to 100 ml of sterile LB broth containing 100 µg/ml ampicillin and 20 µg/ml kanamycin and incubated overnight at 37°C, shaken at 220 rpm. 20 ml of the overnight culture were used to inoculate 1L of sterile LB broth containing 100 µg/ml ampicillin and 20 µg/ml kanamycin and incubated at 37°C, shaken at 220 rpm. Protein expression was induced with 1 mM IPTG (from Melford Laboratories, UK) when the growing culture reached an OD₆₀₀ of 0.7. The culture was further incubated under the above conditions until the OD₆₀₀ reached 1.2. Bacterial cells were pelleted at 8,000 rpm for 30 minutes and resuspended in 5 ml of pre-filtered PBS without NaCl (PB(-S)) (8 mM Na₂HPO₄, 1.5 mM KH₂PO₄, 2.5 mM KCl, pH 7.6). The cell suspension was passed through a cell disrupter (One Shot Model, Constant Systems Ltd, UK) at 12 kPa and then centrifuged at 13 k rpm for 30 minutes to remove cell debris. The supernatant was passed through a 0.2 µm syringe filter (Acrodisc with Supor Membrane, Pall Corporation, USA) and purified soon after.

6.2.3 TOXIN PURIFICATION AND IDENTIFICATION

Crude cell extract was purified on an Applied Biosystems, BioCAD 700E workstation using a hydrophobic interaction column (HIC) with a column volume (CV) of 1.7 ml packed with POROS 20PE matrix with 65% void volume. The column was equilibrated in PBS with 0.7 M NaCl (shortened to 0.7 M PBS) for at least 14 CV at a flow rate between 2.0 and 4.0 CV/min. 5 ml of the extract was injected into the column which was then washed with

double distilled water. The column pressure was kept below 1600 psi at all times and progress was monitored with the built-in spectrophotometer at wavelengths of 280 nm and 260 nm from a quartz cell with a 3 mm path length. Fractions from the wash step were collected and those containing pneumolysin were kept.

Although good purity was achieved with HIC, fractions containing toxin were further purified to remove nucleic acids and any minor contaminants for the interaction studies. The fractions were dialysed against PB(-S) and concentrated using an Amicon stirred ultrafiltration cell (Millipore, USA) cell with a YM10 membrane to a final volume less than 5 ml. An anion exchange column (AEC) with the same CV and packing conditions as the HIC but with POROS 20HQ was equilibrated with PB(-S) for 10 CV at flow rates between 4 to 6 CV/min. The concentrated sample was injected into the column prior to applying a NaCl gradient (from 0 - 1 M NaCl) for 10 minutes with pure toxin eluting at 150 mM NaCl.

Pneumolysin aggregates were removed by SEC. This involved concentrating the purified sample using a ultrafiltration cell and a minicon B-15 concentrator (Amicon) to a final volume of 1ml. This was injected into a 75ml superdex-75 (Pharmacia) size exclusion column (SEC) using the FPLC system from Pharmacia Biotech. The column was equilibrated for 4 hours at a flow rate of 0.5 ml/min with 125 mM PBS except when pneumolysin was being purified for crystallisation purposes, when 25mM tris-HCl at pH 7.5, a commonly used buffer for crystallisation trials was used instead.

A 10% SDS-PAGE gel under denaturing, reducing conditions and a Western blot were used to confirm presence of toxin. Protocols for both techniques are described by Sambrook *et al.* (1989). Concentration of pure toxin was determined from measuring absorbance at 280 nm using an extinction coefficient (ϵ) of $1.36 \text{ cm}^2\text{mg}^{-1}$ (Gilbert *et al.*, 1998). Activity was determined using the haemolytic assay outlined below.

6.2.4 HAEMOLYTIC ASSAY

The 'spot test' haemolytic assay was used to measure the activity of pneumolysin. 50 μ l PBS was pipetted into each well in a 96 well microtitre plate. 50 μ l of diluted toxin was added to the first well, which was used to make further two fold dilutions to neighbouring wells by extracting 50 μ l from the first well and adding it to the next well and so on. The pipette tip was changed for each successive dilution to prevent contamination from residual toxin.

Sheep blood containing heparin obtained from E & O Laboratories, Scotland was washed three times with five volumes of PBS and then spun at 13,000 rpm for a few minutes in a bench-top centrifuge. The pelleted erythrocytes were resuspended in PBS to make a 2% w/v final concentration. 50 μ l of this was added to each well and left to incubate for 30 minutes at 37°C. Erythrocytes that did not lyse sedimented to the bottom of the well where they formed a 'red spot'. The haemolytic unit is the number of dilutions required to cause 50 % cell lysis i.e. for the red spot to shrink to half the size of that resulting from the undiluted toxin. Haemolytic activity was expressed either as haemolytic units per ml (u/ml) or as haemolytic units per mg (u/mg). Because the total volume of suspension in each well was only 100 μ l, the haemolytic unit was factored by 10 to give a value for haemolytic units per ml. For example, if a solution of pure toxin at 0.8 mg/ml requires a 1:16384 dilution to lyse 50% of erythrocytes, then its activity is equivalent to 163840 u/ml or $163840 / 0.8 = 204162$ u/mg.

6.2.5 DERIVATISING PNEUMOLYSIN WITH DTNB (DITHIO(BIS)NITROBENZOATE)

Immediately after pneumolysin had eluted from the SEC column, it was diluted to a concentration of less than 1 μ M in PBS and then derivatised according to a protocol outlined by Gilbert *et al.* (1998) and Wang *et al.* (1996). This involved adding the diluted toxin to an 80-fold molar excess of DTNB dissolved in 100 mM Na-acetate adjusted to pH 5.0 with glacial acetic acid. This was left for at least 1 hour by which time the A_{412} for free TNB⁻ had stabilised. The toxin DTNB mixture was dialysed against a 200-fold volume excess of 150 mM deuterated PBS in a sealed container for Neutron scattering experiments. For crystallisation trials, the mixture was dialysed with a 2000-fold volume excess of Tris-HCl or with 150 mM PBS for other experiments. Samples were dialysed for over 70 hours at 4°C under stirred conditions with at least five changes to the dialysate. Samples were stored at -20°C and when required were spun for a few minutes at 13,000 rpm on a bench-top centrifuge followed by concentrating the supernatant using a B15 minicon concentrator (Amicon).

6.2.6 DETERMINING SAMPLE HOMOGENEITY

Sedimentation equilibrium (SE) runs using the Optima XL-A analytical ultracentrifuge (Beckman) were used to check the homogeneity of the sample. Because SE can determine the absolute molecular weight of a molecule, no instrument calibration or shape assumptions are required. By measuring the average molecular weight of pneumolysin at different

concentrations, the presence of oligomers would manifest itself by raising the average molecular weight of the toxin at infinite dilution.

100 μl aliquots of fractions collected after AEC and SEC were made to concentrations ranging from 8.5 μM to 1.8 μM and 1.6 μM to 0.4 μM respectively. 80 μl from each of these aliquots was loaded into multichannel centrepieces (a total of three centrepieces were used to run the samples) containing a total of six cells divided into two symmetric columns. Sample and buffer were loaded into their respective cells, and spun at 3 k rpm, 16 k rpm, 20 k rpm and 40 k rpm, in this particular order. The concentration of pneumolysin in each sample was checked at 3 k rpm from its absorbance readings and was monitored for a short period to ensure that there was no leakage. The centrifuge speed was increased to 16 k rpm and once equilibrium had been established was increased to 20 k rpm. Equilibrium was checked at both speeds as a precautionary method that would ensure an even concentration distribution. Scans were taken periodically at both speeds for approximately 20 hours by which time the absorbance readings remained constant indicating that equilibrium had been reached. The speed was then increased further to 40 k rpm, with two scans taken after a period of four hours to measure the baseline. This was subtracted from the equilibrium measurements taken at 16 k rpm and 20 k rpm. The similarity of the trace from both scans indicated that all large particles had been pelleted. The absorbance reading from this scan was used to obtain a baseline calibration for the experiment.

Data were fitted to the ASSOC4 model from the Beckman *Origin* software according to the equation given by Kim *et al.* (1977) cited in (Gilbert, 1998).

6.3 RESULTS AND DISCUSSION

6.3.1 PRODUCTION AND PURIFICATION OF PNEUMOLYSIN

The results presented in this section are based on a specific example that is representative of findings found on numerous, separate occasions. It charts the progress from bacterial growth, where the toxin was expressed intracellularly, followed by the extraction and purification of the toxin. The first 280 minutes of growth for this particular strain of *E. coli* (Figure 6.3) shows that the lag phase takes less than 100 minutes. The culture was induced with IPTG at mid exponential phase (approximately 3 hours after seeding the culture) and bacteria harvested when the absorbance measurement at 600 nm reached 1.2 Au.

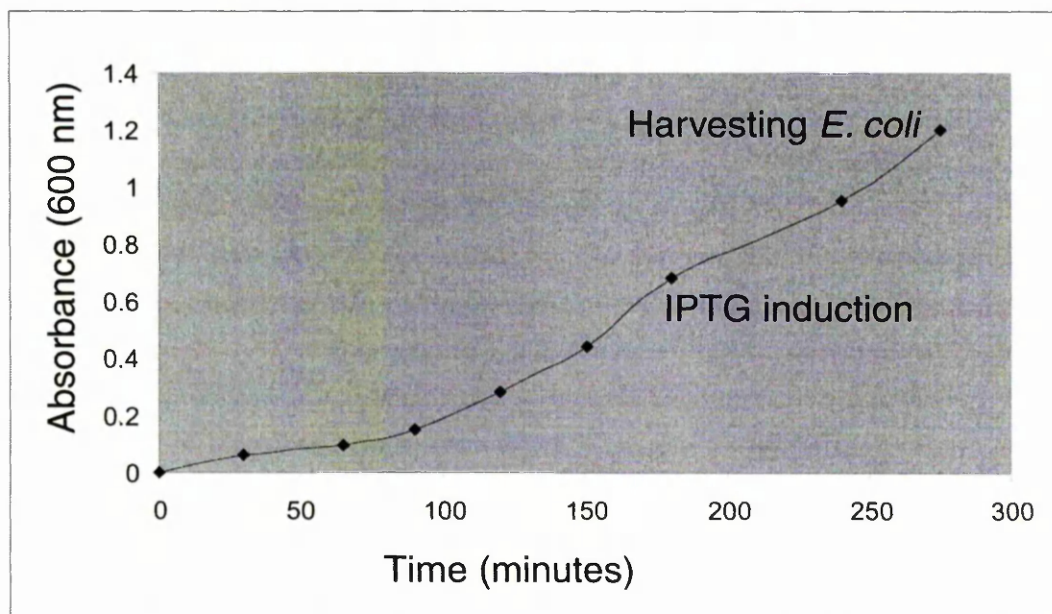


Figure 6.3 The first 280 minutes of growth for the *E. coli* strain M15 carrying the pkkwt plasmid. Growth was monitored by taking periodic spectroscopic measurements at 600 nm.

During the exponential growth phase, the decrease in growth rate after induction may have been caused by a combination of the following factors: redirection of energy utilisation from growth to toxin production; accumulation of intracellular toxin having harmful effects on the bacterium; laboratory handling during the induction step may have destabilised the constancy of the growth conditions. However the effects were mild and stationary phase was usually encountered after 320 minutes by which time the cells had been harvested by centrifugation at 8,000 rpm.

Having disrupted the cells, the filtered cell extract which was in 0.7 M PBS was passed through a perfusion chromatography column containing POROS 20PE media. The majority of the extract passed directly through the column as part of the wash-through, with pneumolysin coming off isocratically in the shoulder peak giving relatively pure pneumolysin fractions as shown by the gel in Figure 6.4A. This was confirmed by a Western blot of those fractions as shown in Figure 6.4B. The presence of slight degradation products in fractions 5 and 6 (F5 and F6) was probably caused by proteases native to *E. coli* (Baneyx, 1999). These bands showing degradation were not visible in the scanned image and so marks have been added to indicate both position and intensity (relative to the main band) of the degraded bands.

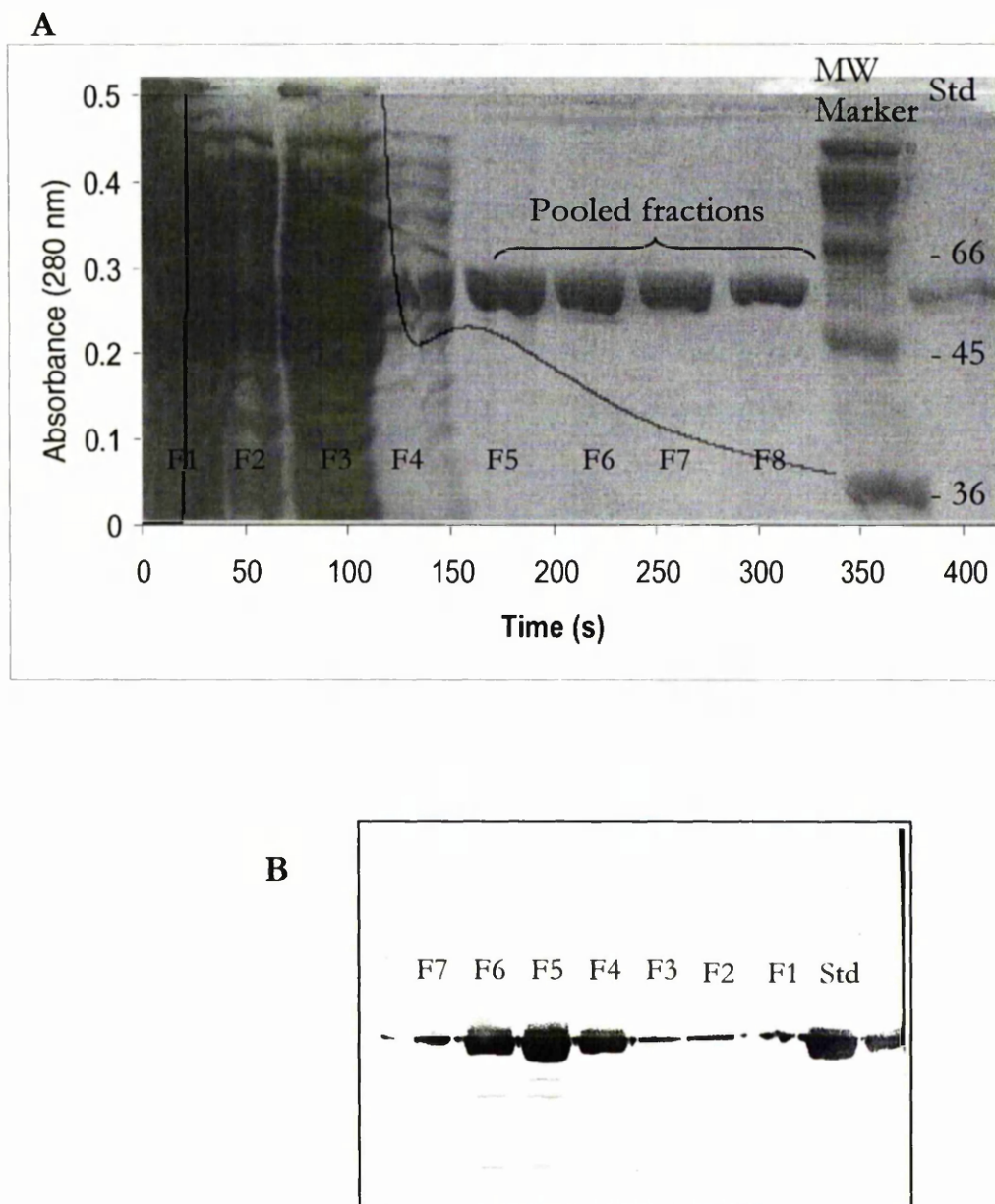


Figure 6.4 Fractions F1 to F8 from the HIC purification were run on 10% SDS PAGE (**A**). The gel is superimposed on the purification chromatogram monitored at 280 nm. The appearance of a shoulder peak 150 s after injection signifies the presence of pneumolysin; as indicated by the bands from F5 to F8, which are correctly aligned with the pneumolysin standard (Std) and correspond to 53 kDa (according to the molecular weight (MW) marker, values for which are given in kDa). This was also confirmed by a Western blot of those fractions (**B**).

The presence of bands in fractions 1, 2 and 3 of the Western blot (F1, F2 and F3) indicates that not all pneumolysin binds to the column. The capacity of the column according to the product guide (Media Selection Guide, 1996) is 12 mg. The crude extract is estimated to contain 6 mg of toxin per litre of broth (see section 6.3.2). Therefore the column should

theoretically be capable of binding all the toxin. Possible causes for the presence of toxin in the wash through could be that other proteins in the extract cause 'crowding effects' that can impede diffusion of toxin towards the ligand (Minton, 1989). Diluting the extract in a volume larger than the stated 5 ml of PB(-S) did not improve matters significantly. Therefore it is more likely that pressure (approximately 1500 psi) and salt concentration inside the column may have induced the toxin to aggregate, which in turn will have reduced its affinity for the ligand, causing it come out in the wash through.

Fractions F5 to F8 had a relatively high absorbance at 260 nm, with an $A_{280}:A_{260}$ of approximately 1:1 and at 320 nm with an $A_{280}:A_{320}$ of approximately 4:1. These readings were indicative of the presence of nucleic acids and aggregates, respectively. An anion exchange step was used to remove nucleic acids and any other minor contaminants such as degradation products. The 20 HQ POROS column with a linear (0 to 0.5 M) NaCl gradient rising linearly beyond 0.5 M was used to elute pneumolysin at 150 mM NaCl (Figure 6.5). The $A_{280}:A_{260}$ ratio for collected fractions was approximately 2:1 and for $A_{280}:A_{320}$, it was approximately 10:1. Both ratio's indicate that the AEC step has substantially reduced the presence of both nucleic acids and aggregates.

Contaminants from the 20 HQ column were found to elute either immediately after the sample had been injected into the column or when the NaCl concentration increased beyond 200 mM NaCl (Figure 6.6). By removing nucleic acids from the HIC fractions hence removing the slight absorbance that occurs at 280 nm, the toxin concentration could be quantified from its A_{280} reading more accurately. Purification from a 1 L growth culture using both HIC and AEC POROS columns could yield up to 4 mg of toxin, which is similar to that obtained by R. Gilbert (personal communication).

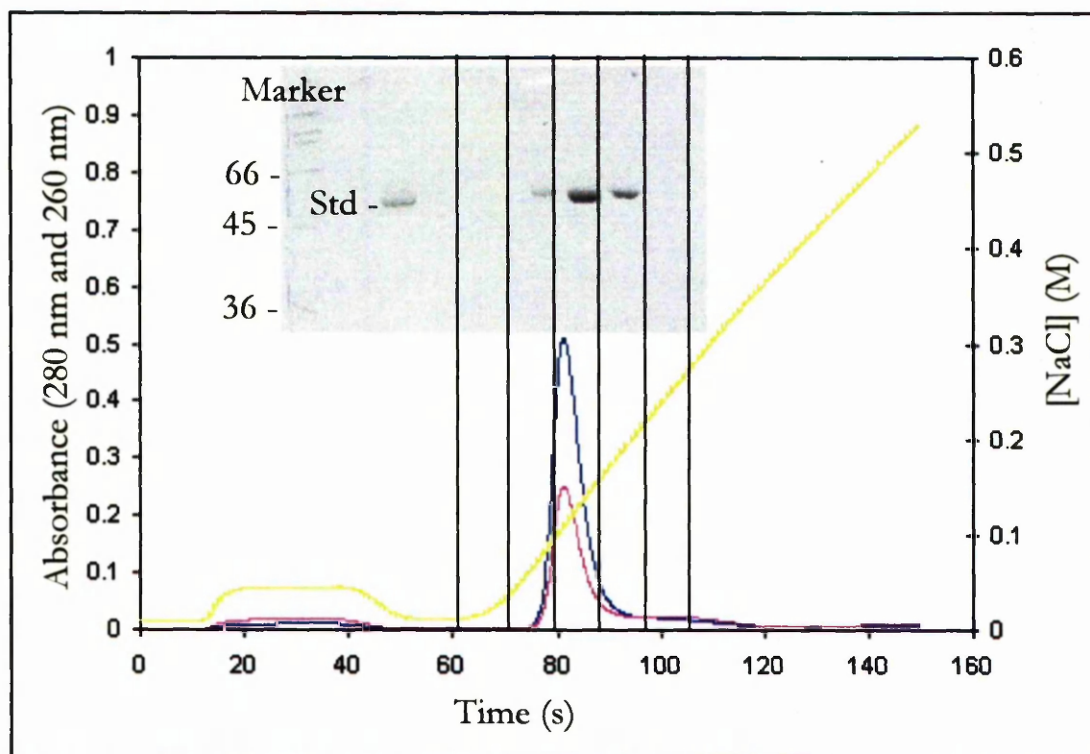


Figure 6.5 Pooled fractions (F5 to F8) from HIC were dialysed in PB(-S) and concentrated prior to injecting onto an AEC column. The chromatogram shows absorbance at 280 nm (—) and 260 nm (—), and the NaCl gradient (—) up to 0.5 M. Pneumolysin elutes after 70 s as a single distinct peak collected in three fractions showing clean single bands that correspond with pneumolysin standard and 5 kDa molecular weight marker.

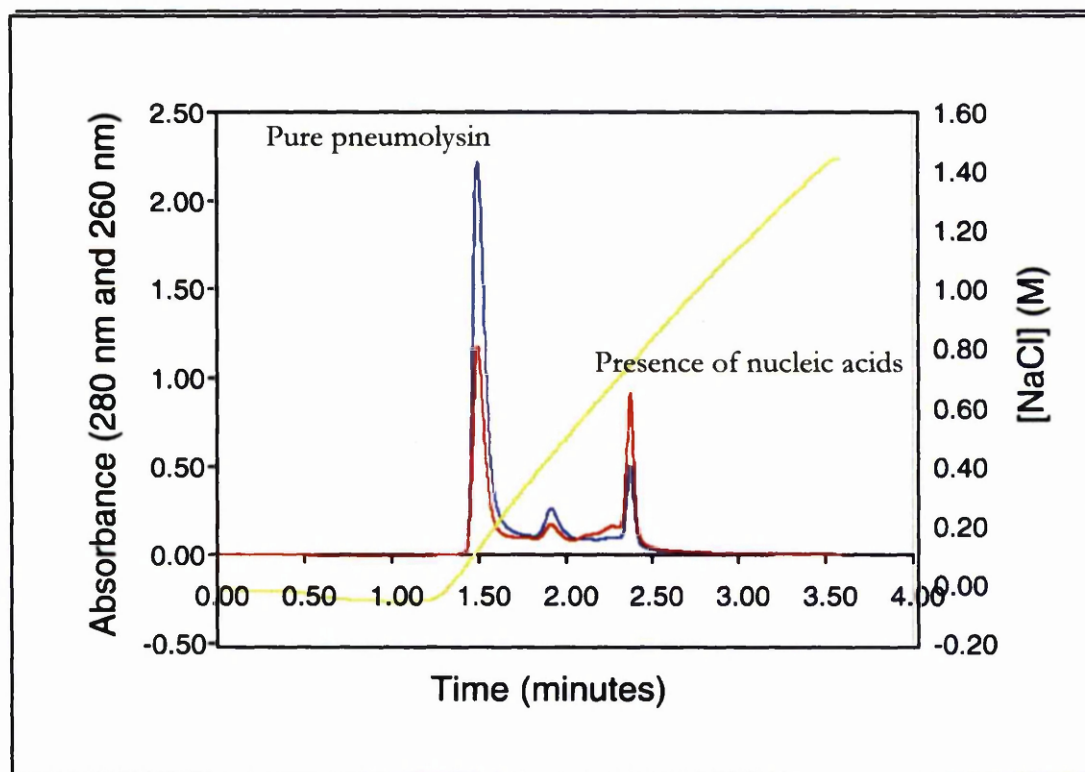


Figure 6.6 Elution of contaminants (e.g. nucleic acids) from an AEC column at higher salt concentrations. Chromatogram shows absorbance at 280 nm (—) and 260 nm (—), and NaCl gradient (—) up to 1.5 M.

Although contaminants may not always be present after HIC purification, the application of a second orthogonal purification step is recommended as a precautionary step by purification guide books (Sheehan, 2000). Comparison of the absorbance spectra between post HIC and post AEC indicates that both nucleic acids and aggregates have been removed by the latter step. Therefore purification is only complete after the AEC step.

6.3.2 EFFICIENCY OF PURIFICATION

The toxin yield was determined at five different stages to determine the efficiency of the production and purification process for a 1 L growth culture. These stages were (i) before inducing the *lac* promoter with IPTG; (ii) at stationary growth of the bacterial culture (prior to disrupting bacterial cells); (iii) after bacterial cell disruption; (iv) after HIC and (v) after AEC purification. Toxin yield was calculated from the haemolytic activity of samples taken from each stage. For the two stages prior to cell disruption, sonication of 10 ml cell suspension was used to lyse cells. The haemolytic activity for pure toxin ranged between 10^5 to 10^6 u/ml which was within an order of that obtained by Mitchell *et al.* (1989), which was 10^6 u/ml.

STAGES	Yield from each production batch (mg)				
	1	2	3	4	5
Pre-IPTG induction	0.2	2.5	1.1	0.9	1.3
Pre-cell harvesting	26.25	16	13	7	25
Post cell disruption	8.4	8	4.4	4.5	4
Post HIC separation	10	4	5	3.4	6
Post AEC separation	4.25	2.5	2.15	2.25	4

Table 6.1 Pneumolysin yield from a 1 L growth culture at different stages of expression and purification as determined by haemolytic activity measurements.

The amount of pneumolysin present at each stage was calculated making the assumption that haemolytic activity is proportional to toxin concentration. Therefore the ratio of the haemolytic activity at a particular stage (*HS*), and for pure toxin (*HP*) is proportional to the ratio of toxin concentration at that stage (*CS*) and for pure pneumolysin (*CP*) i.e. $HS/HP \propto CS/CP$. By determining *CS* in this way, it was possible to determine the yield at each stage as shown in Table 6.1. The results show that pneumolysin was being expressed prior to addition of IPTG even though expression was increased dramatically upon induction. The expected fall in toxin yield between the cell harvesting stage and the final purification stage did occur, but batches 1, 3 and 5 showed a slight increase in yield between cell disruption and HIC purification. The likely cause of this may have been experimental error in terms of determining the precise dilution at which 50% haemolysis had occurred. Alternatively haemolytic activity as discussed in the following chapter (chapter 7) is known to vary depending upon the content of specific cations (such as calcium) in solution (Korchev *et al.*, 1998) or due to a decrease in the effective pneumolysin concentration. The latter can arise if pneumolysin monomers oligomerise in the presence of free cholesterol (Morgan *et al.*, 1995) released from cell lysate or due to the non-specific adsorption of toxin to artificial vesicles formed by the mechanical cell disruption process.

The values from Table 6.1 when plotted on a graph (Figure 6.7) showed that the biggest variation in yield stemmed from the cell harvesting stage. The reason is again most likely to be caused by attempting to accurately identify the well in which 50% lysis has occurred. So if either of the neighbouring wells were selected, the final yield would be either half or double the original value.

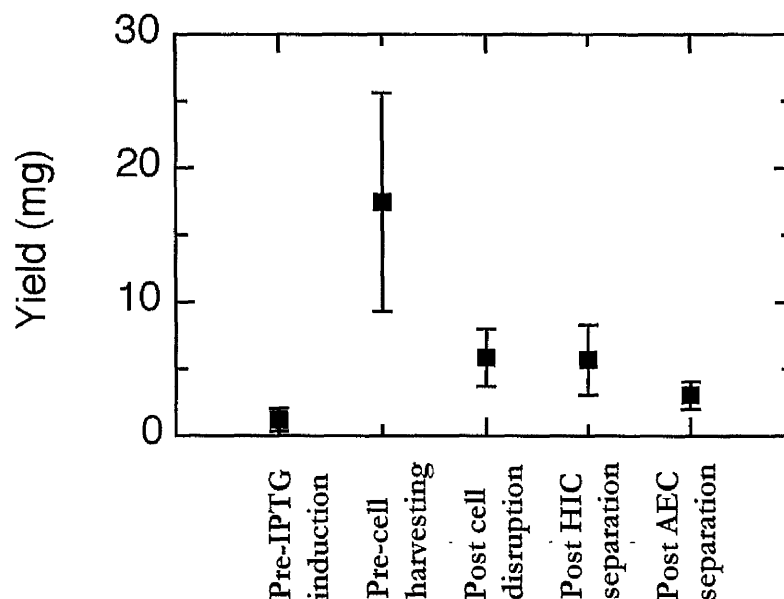


Figure 6.7 Toxin yield at different stages of the production and purification procedure.

The drop in toxin yield from the HIC purification step to AEC is most likely to be due to the formation of pneumolysin aggregates caused by the high ionic strength of the crude extract (as mentioned in the previous section). Pure pneumolysin (in 150 mM PBS) when added to 1 M PBS did show visible signs of precipitation after overnight storage at 4 °C. The same was found to occur when pneumolysin in 0.7 M PBS was concentrated and then diluted with PB(-S). These observations suggest that changing the ionic strength of the solvent can induce aggregation. Binding studies on α -haemolysin, which is also susceptible to aggregation at high ionic strength and by high pH (6-8), to lipid vesicles showed that as the aggregate size grew bigger, the amount of toxin binding to the vesicles dropped (Ostolaza *et al.*, 1997). Similarly, if aggregates are formed in the crude extract as the result of high salt concentrations or by the presence of lipids in the crude extract, its size will lower its binding affinity to the ligands which would consequentially result in the loss of toxin in the wash through (Figure 6.4).

6.3.3 ALTERNATIVE STRATEGIES FOR OPTIMISING PURIFICATION

Although the toxin had been successfully purified from crude extract, preliminary investigations were conducted on the feasibility of modifying the purification strategy. The initial aim was to improve the binding of toxin to the hydrophobic column. Three different strategies were attempted:

1. Equilibrating the column at higher NaCl concentrations

2. Replacing NaCl with $(\text{NH}_4)_2\text{SO}_4$
3. Using a stronger hydrophobic matrix

(1) Because pneumolysin interacts weakly with the PE matrix at 0.7 M NaCl, a higher concentration of NaCl was used to bind the toxin to the matrix. Pneumolysin bound to a column pre-equilibrated with 1.0 M PBS. But when this column was washed with PB(-S), two proteins (one of which was pneumolysin) eluted. Figure 6.8 shows an extreme example of this case, showing two very distinct protein bands. A Western blot of the fractions containing the double bands indicated that only the higher molecular weight band was pneumolysin. The proteins could not be separated from each other even when a shallow elution gradient (approximately 10 minutes at 2 CV/min) from 1.0 M PBS to PB(-S) was applied to the HIC column. Similarly an AEC step was unable to separate the two bands. This suggests that a protein of similar molecular weight is also present in the purified fractions, although there may be the remote possibility that it may be pneumolysin. For instance, pneumolysin may undergo a conformational change at high ionic strengths (> 0.7 M NaCl) which may make it more susceptible to degradation by endogenous proteases. This would explain the presence of low molecular weight fragments. α -haemolysin has also been shown to undergo a conformational change at pH 7-8, which has been shown by changes in its fluorescence emission intensity and its difference in susceptibility to proteolytic cleavage by papain (Ostolaza *et al.*, 1997).

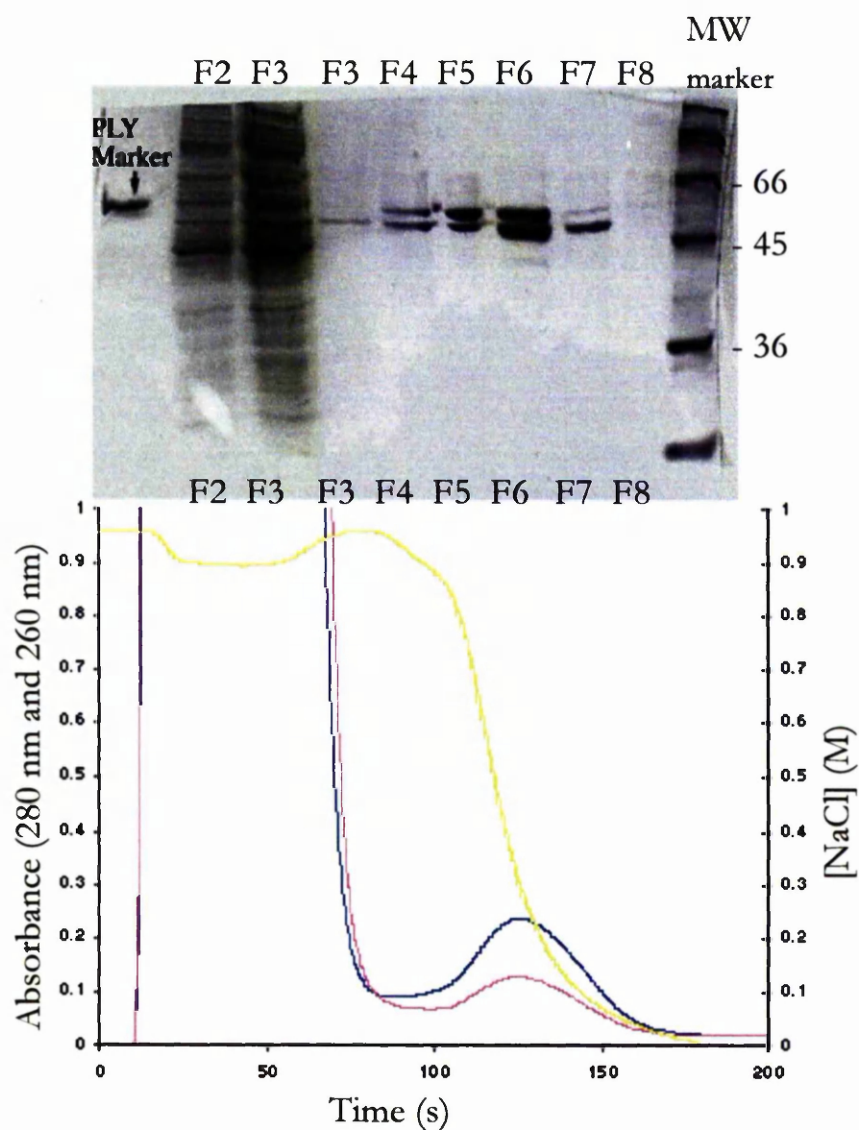


Figure 6.8 A chromatogram of HIC purification of pneumolysin from crude extract is aligned with a 10% SDS PAGE gel showing protein bands from each fraction (F2 to F8), the pneumolysin (ply) marker and molecular weight (MW) standards (values for which are given in kDa). The column was equilibrated with 1.0 M PBS and set at a flow rate of 4 CV/min. Toxin eluted with another protein when the concentration of NaCl fell below 0.8 M PBS.

(2) Ammonium sulphate is more lyotropic than NaCl and should therefore be more suitable for binding proteins to the hydrophobic matrix. The 20PE column when equilibrated with 250mM and 400 mM $(\text{NH}_4)_2\text{SO}_4$ did not bind pneumolysin but instead retarded its progress down the column. Although $(\text{NH}_4)_2\text{SO}_4$ concentrations tested were comparatively lower than

NaCl, the lyotropic nature of the $(\text{NH}_4)_2\text{SO}_4$ salt did not appear to show significant benefits for binding the toxin to the column at low salt concentrations.

(3) A stronger matrix (HP2) containing a higher density of phenyl ether ligands was also used to purify the toxin without success. Crude extract injected into a column equilibrated in 0.6 M PBS resulted in a single wash-through, which tested positively for haemolytic activity indicating that the toxin did not bind to the column. 5 ml of pre-purified pneumolysin at a concentration of 0.5 mg/ml was also passed through the column to check if the pneumolysin could bind to the matrix. The absence of an A_{280} when the column was equilibrated in 0.6, 0.2 and no salt implied that pneumolysin had bound to the column. The lack of haemolytic activity in the collected fractions indicated that the toxin had bound to the column. This seems to suggest that toxin present in crude extract binds poorly to the column where as with no extract, it binds tightly to the column. One suggestion is that lipopolysaccharide (LPS) from *E. coli* present in the extract may have coated the column preventing the toxin from binding to it. Where as with pure pneumolysin the absence of LPS and other hydrophobic molecules allowed the toxin to bind tightly to the column and was only released when washed with 0.6 M NaOH and strong chaotropic agents such as guanidium thiocyanate.

One of the features of the BIOCAD 700E workstation is that two columns can be run in tandem so that both HIC and AEC steps could be completed in a single run. The BIOCAD was set-up so that crude extract when injected passed through the HIC (20PE) column, which was equilibrated with 0.6 M PBS, with two CV of eluent (directly after sample injection) passing to waste. The AEC (20HQ) column, which was equilibrated in PB(-S) was then brought in-line. This should have ensured that the initial wash through did not pass through the AEC column. Results from this purification revealed fractions containing a lot of impurities coming through from remnants of the wash through. Purer fractions were obtained if the discarded volume was increased from 2 to 6 CV, but yielding much lower levels of toxin. The set-up showed that it was possible to purify pneumolysin in a single run but with an inevitable loss of toxin (approximately 40% efficiency when compared with a single column purification).

Initial attempts using the AEC as a first step gave no distinct chromatographic peaks. The method used ran at 4 CV/minute, for 10 minutes with a salt gradient ranging from PB(-S) to 1.0 M PBS. Fractions were not assayed for haemolytic activity. Although further work is necessary to optimise this procedure it may be prove to be a worthwhile avenue for further investigation.

In summary, it was possible to obtain pneumolysin using perfusion chromatography at yields that were generally greater than 2 mg/ml, which was comparable to similar protocols (Gilbert

et al., 1998; Mitchell *et al.*, 1989) using the matrix provided by Pharmacia. The flow rate used for the HIC (4 CV/min) was probably the maximum attainable without compromising either the resolution or yield, whereas with the AEC faster rates (up to 8 CV/min) had been used. The faster flow rate was not used as part of the standard procedure because it raised the pressure inside the column close to its limit.

6.3.4 CHEMICAL DERIVATISATION

Electron micrographs, sedimentation equilibrium, sedimentation velocity and small angle neutron scattering have shown pneumolysin to oligomerise in deuterated solutions. The extent of oligomerisation can be reduced by chemical derivatisation with DTNB (Gilbert *et al.*, 1998). Protein-TNB conjugates have a higher absorbance at 337 nm than underivatised protein owing to the specific absorbance by TNB. The absorbance ratio for derivatised toxin taken at 278 nm and 337 nm can be used as an indicator of how much sample has been derivatised. Wang *et al.* (1996) in a separate study involving thioredoxin reductase obtained a ratio of 13:5, which indicated that the sample had been completely derivatised. This value was much lower than the 10:1 value measured by Gilbert *et al.* (1998) for pneumolysin. Reasons for the presence of underivatised toxin were not suggested by Gibert and co-workers, but one possibility could be that the underivatised toxin represented toxin oligomers that would have been present prior to derivatisation and thus could not be chemically modified.

Evidence of pneumolysin oligomers was obtained by sedimentation velocity studies performed by Gilbert *et al.* (1998), indicating the presence of a second high molecular weight species, considered to be the toxin dimer, and also from this present study, where aggregates were occasionally found to elute from the SEC column early in purification. This is illustrated by the SEC chromatogram in Figure 6.9. The two absorbance peaks on the chromatogram indicate two differently sized proteins, the latter being smaller in size but present in considerably larger amounts. Besides being larger in size, evidence to suggest that the first peak was indeed pneumolysin aggregates came from two further studies of the SEC eluent: SDS-PAGE and a haemolytic test. Samples taken from both peaks when run on a 10% SDS-PAGE gave a single band for both samples, which ran with the same electrophoretic mobility as the pneumolysin standard. Haemolytic activity was detected only for the sample collected from the second peak.

Aggregation was particularly noticeable when toxin samples greater than 3 mg/ml were derivatised. They were detected either visually, from the cloudy appearance of the sample or from the size of the pellet that resulted after centrifuging the sample for a few minutes at 13,000 rpm on a bench top centrifuge. It tended to be noticed an hour after glacial acetic acid

(pH 5) had been added and was therefore presumed to be the most likely cause. One explanation for this behaviour is that the electric charge surrounding pneumolysin which has its isoelectric point (pI) at 5.05 (as determined by *PROMOLP*, mentioned in (Hantgan *et al.*, 1999)) becomes polarised as the pH is lowered. This would then result in electrostatic interactions between neighbouring monomers. By ensuring that the toxin was diluted to less than 1 μM (a step not present in the original protocol described by Gilbert *et al.* (1998)), the volume of the DTNB solution that needed to be added to the toxin was much less, and hence the drop in pH was also less. Diluting the toxin sample also reduced the likelihood of any protein-protein interactions. The end result was that there was no visible signs of aggregation and an $A_{280} : A_{337}$ was found to be approximately 4:1, a huge improvement on the 10:1 reported by Gilbert *et al.* (1998) but still not quite as low as the ratio obtained by Wang *et al.* (1996).

To check whether this modification to the protocol did actually improve the efficiency of derivatisation, 2 mg of pneumolysin, prior to and after derivatisation (with an $A_{280} : A_{337}$ of 7:2) were sent to Dr M. Scott, Department of Chemistry, University of Edinburgh for mass spectrometric analysis with a Platform II Mass Spectrometer running in electrospray positive ion mode. Readings taken for the underivatised sample (Figure 6.10A) showed a peak at 52,771 Da when deconvoluted, which is close to the mass of pneumolysin as determined from its sequence (52,772 Da) (Walker *et al.*, 1987). The calculated mass of pneumolysin-TNB was 52,970 Da, which was within the error limit of 52,967 Da as determined by mass spectrometry (Figure 6.10B). Finally a 1:1 mixture for both pneumolysin and pneumolysin-TNB gave two peaks that corresponded to the masses of each component (Figure 6.10C). The mass spectrometry measurements suggest that the derivatisation process went to completion. As the absorbance ratio for the derivatised sample was 7:2, this ratio presumably indicates complete derivatisation for the specific case of pneumolysin.

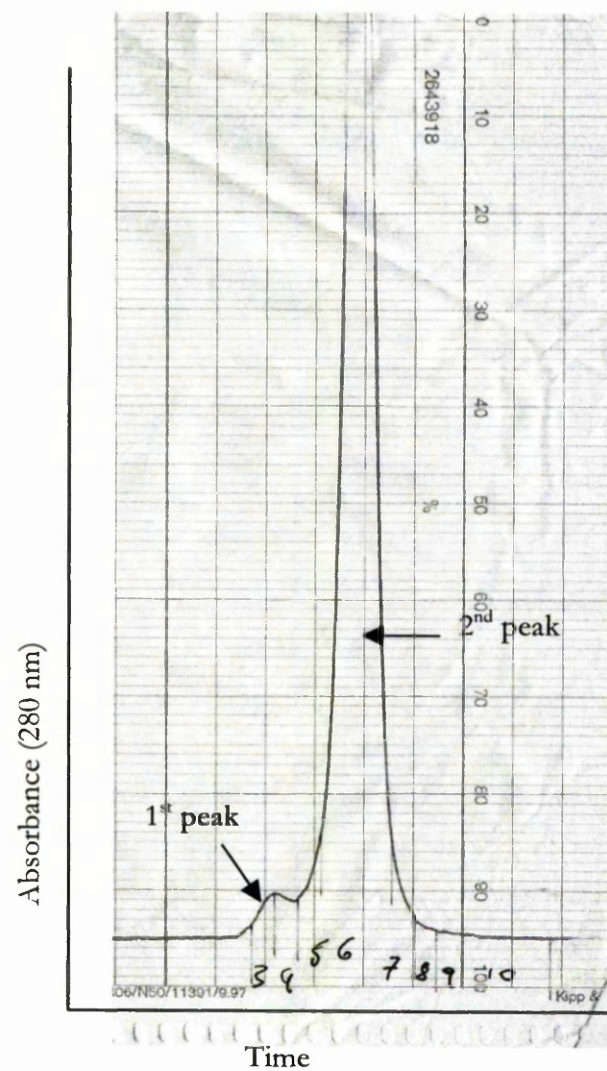
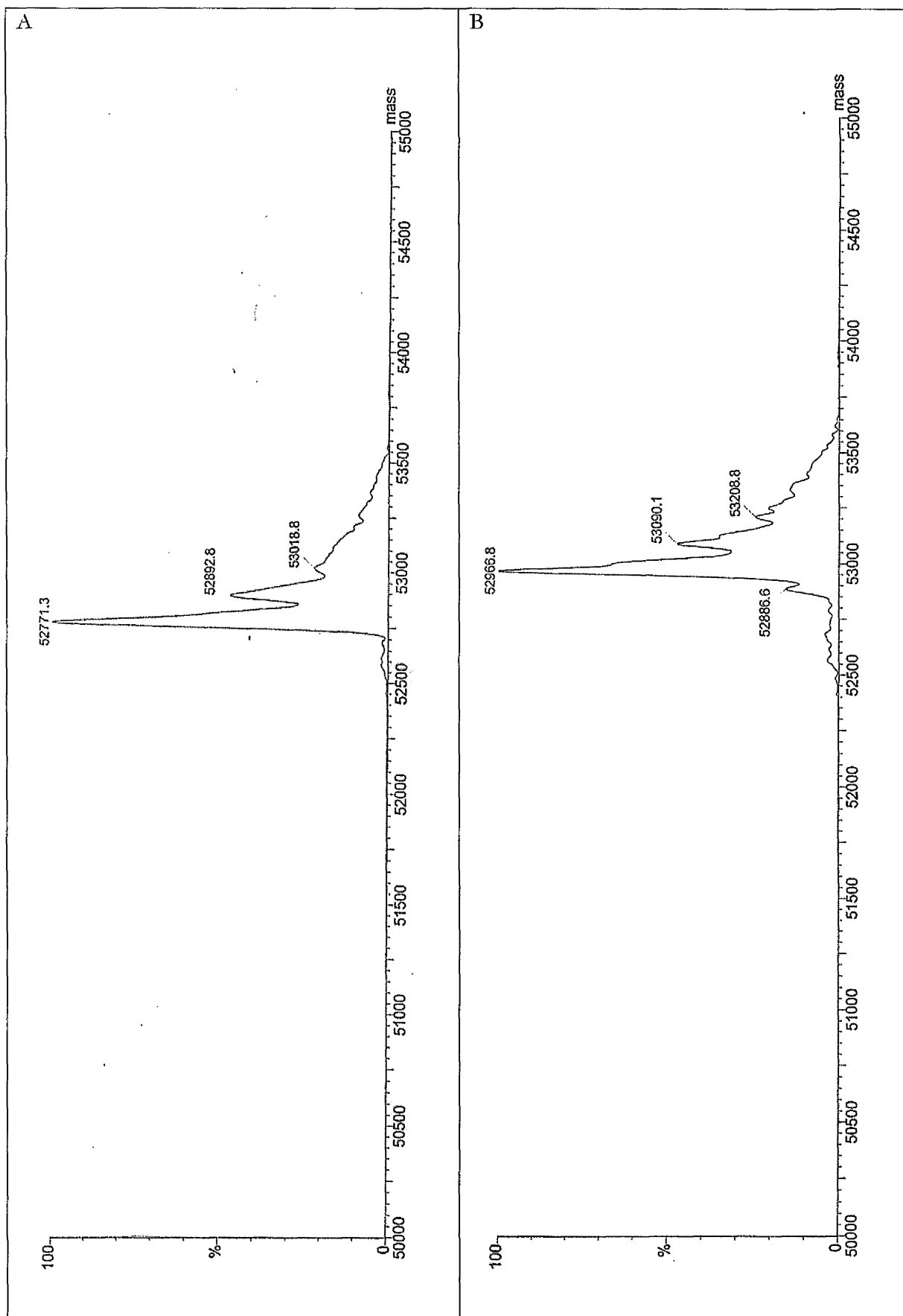


Figure 6.9 SEC chromatogram showing two peaks in absorbance at 280 nm. The first peak corresponds to the elution of pneumolysin aggregates whereas the second peak corresponds to monomeric pneumolysin.



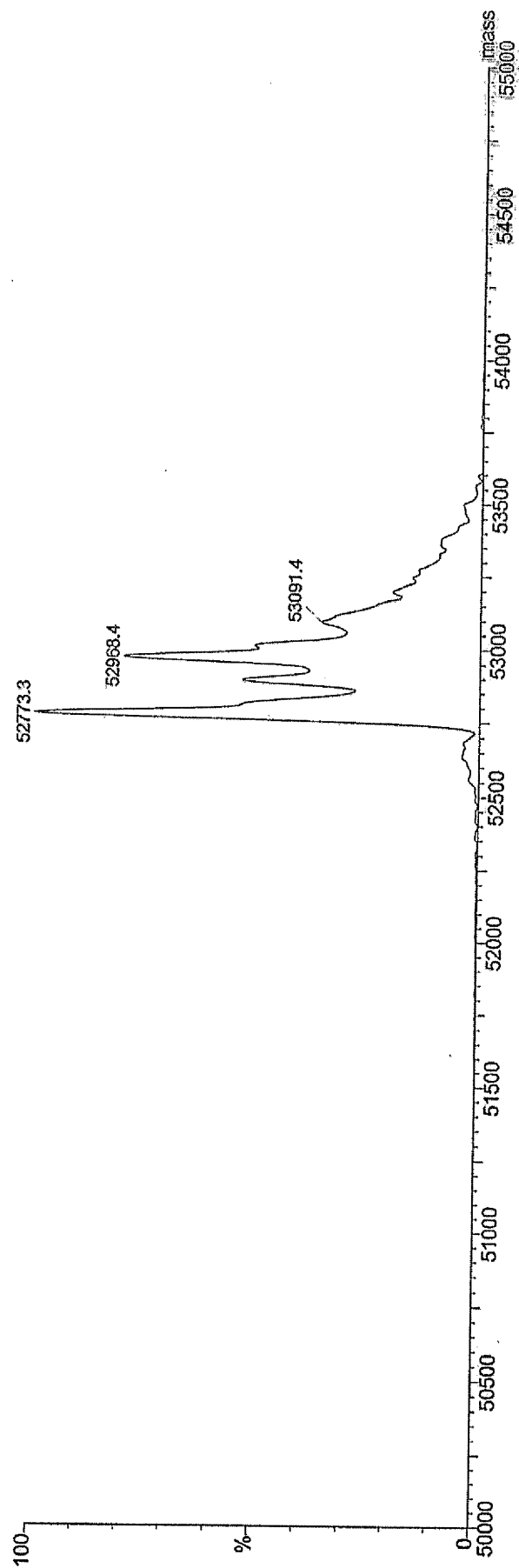


Figure 6.10 Mass spectrometry traces for (A) pneumolysin, (B) pneumolysin-TNB conjugate, and (C) a 1:1 mixture of pneumolysin and pneumolysin-TNB.

As eluded to previously, not all samples after AEC contained aggregates. In many cases, SEC would only detect a single peak. An SE run on pneumolysin taken before SEC showed a remarkably greater degree of fluctuation (between 42 kDa and 53 kDa), whereas after SEC the molecular weight average of the toxin taken at infinite dilution, settled to between 45 kDa and 48 kDa (Figure 6.11). This value was lower than that typically obtained (> 50 kDa) from previous hydrodynamic studies (Gilbert *et al.*, 1998). The apparent rise in molecular weight with concentration was indicative of the toxins tendency to self associate (Gilbert *et al.*, 1998). But because the SE runs determined a molecular weight average of approximately 47 kDa and because only a single peak was detected in the SEC chromatogram, it was unlikely that aggregates were present in the sample.

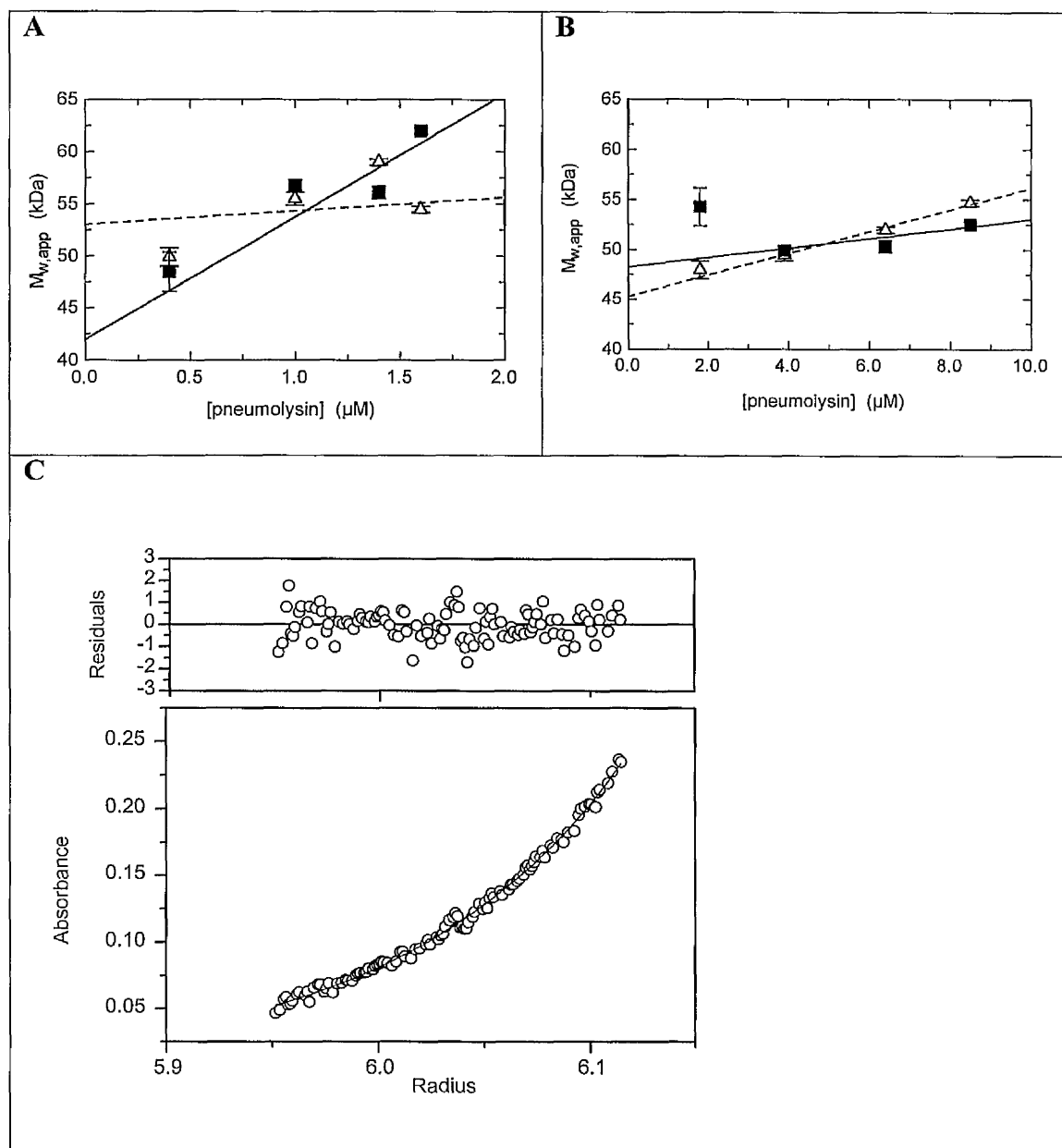


Figure 6.11 SE studies for pneumolysin at 16 k (Δ) fitted with a dashed line and at 20 k (\blacksquare) fitted with a solid line, before (A) and after (B) SEC. (C) Fitting of experimental data (0.4 μM at 16 k) showing concentration radial distribution with minimal residual deviation.

6.4 CONCLUSION

The initial objective of this study was to assess the practicality of the perfusion chromatography technique as a standard method for purifying pneumolysin from crude extract and to improve the actual chemical derivatisation of TNB to pneumolysin. The results

show that perfusion chromatography using both hydrophobic and anion exchange matrix as a viable method for purifying pneumolysin. A 1 L culture of the *E. coli* strain used during the purification procedure yielded on average 4 mg of pure pneumolysin. Although this yield is comparable to that obtained by Gilbert, the actual time required for purifying a 5 L culture using both HIC and AEC steps can be done in under 3 hours. Although the toxin yield measurements at each of the different production and purification stages were not very accurate (Figure 6.7), it did appear to indicate that significant loss of toxin occurred during the cell harvesting and disruption stages. Some of this loss may have been caused by the mechanical cell disruptor. The disruptor was used to lyse bacterial cells by forcibly squeezing them through a very narrow orifice. But this very process is also likely to result in the formation of lipid vesicles. It is possible that such vesicles will contain pneumolysin, thus preventing it from binding to the chromatographic matrix. Another possibility is that pneumolysin has associated with the membrane components of the cell lysate and as such will have been discarded at an earlier stage of purification. The chemical derivatisation of pneumolysin with TNB was found to improve when performed after SEC and when using dilute toxin concentrations. Evidence for this was shown from mass spectrometry and its $A_{280}:A_{337}$ ratio.

CHAPTER 7

AN ELECTROCHEMICAL METHOD TO STUDY THE LEAKAGE EVENTS IN PNEUMOLYSIN-INDUCED HAEMOLYSIS

7.1 INTRODUCTION

Bacterial protein toxins are good examples of proteins that are both soluble in water and yet capable of inserting into the hydrophobic environment of certain types of membrane. Pneumolysin is thought to induce cell lysis by the following sequence of events (Boulnois *et al.*, 1991): (i) interaction of monomeric toxin with cholesterol (which is found in the membranes of most eukaryotes); (ii) temperature dependent insertion into the membrane bilayer; (iii) oligomerisation of monomers to form pores that can be upto 30 - 45 nm in diameter with a ring width that is 6.5 nm thick (Morgan *et al.*, 1994; Morgan *et al.*, 1995). The movement of water and ions through pores collapses the potential across the membrane, and for certain cells such as erythrocytes, it can result in its lysis (Boulnois *et al.*, 1991). However studies performed by Duncan (1974), Blumenthal and Habig (1984) indicated that lysis occurred immediately after the toxin inserted into the membrane. For non-erythroid cells, cell death is more likely to be caused by the leakage of large molecules such as phosphorylated metabolites (sugar, phosphates, nucleotides, etc.) and cytoplasmic proteins from the cell (Bashford *et al.*, 1985).

7.1.1 ION LEAKAGE FROM ERYTHROCYTES

A variety of different methods using electrochemical techniques have been developed to monitor the movement of ions across membrane channels. They have traditionally used a combination of biological and synthetic membrane systems. In the latter case, track-etched PETP (polyethyleneterephthalate) membranes (Rostovtseva *et al.*, 1996), liposomes and artificial lipid bilayers were generally used to simplify (Bashford *et al.*, 1996) and mimic features present in biological systems such as erythrocytes. However the complexity of biological systems not only makes interpreting data a little more complicated, but it can also give results that disagree with measurements conducted on artificial systems. For example, Alder *et al.* (1990) had found that the size of pores formed by diphtheria toxin differed in natural and artificial membranes. Therefore it is important that studies are carried on both types of systems.

Erythrocytes are commonly used as a model for studying eukaryotic membranes. The human erythrocyte membrane like other mammalian erythrocyte membranes contains four major phospholipid classes distributed unevenly on the inner and outer surface of the membrane. The choline containing phospholipids, sphingomyelin (SM) and phosphatidylcholine (PC) are mostly found on the outer surface of membranes, whereas the aminophospholipids, phosphatidylethanolamine (PE) and phosphatidylserine (PS) are typically concentrated on the cytoplasmic side (Zwaal *et al.*, 1975). In human erythrocytes, approximately 82% of SM and 76% of PC are located on the outer surface and 80% of PE and virtually all of PS on the inner side (Butikofer *et al.*, 1990; Gascard *et al.*, 1991) although there has been some dispute whether PS is also found on the outer surface (Devaux, 1992).

Several factors account for this non-random distribution such as diffusion (passive and facilitated); a pH gradient and active transport (Devaux, 1990). The latter is thought to emerge from either a single or a group of aminophospholipid-specific translocases, which are thought to translocate PE and PS across the membrane in both directions (Devaux, 1992). This enzyme is known to be ATP and temperature dependent (Seigneuret and Devaux, 1984) and requires the presence of 2 mM magnesium and a low intracellular calcium concentration. For instance, 1 μ M calcium inside the cell inhibits the inward translocation of both aminophospholipids, whereas extracellularly, calcium along with other divalent cations will have no effect on translocation at concentrations up to 1 mM (Bibitot *et al.*, 1987). The inhibition of pneumolysin activity that Johnson (1972) encountered upon addition of high levels of divalent ions is likely to be caused by a number of factors that probably includes redistribution of phospholipids in the membrane and pore closure as a result of chelation of divalent ions to negatively charged groups to the external side of the membrane (Bashford *et al.*, 1989). However it is unlikely to be involved in directly preventing the toxin from binding to the cell (Korchev *et al.*, 1998).

7.1.2 THE ENZYMOMETER

The enzymometer is a conductimetric instrument developed by Dr A. Lawrence at the University of Glasgow, which has been used as a teaching tool to study the progress of enzymatic reactions (Lawrence *et al.*, 1998). It was designed to monitor the progress of simple enzymatic reactions, where alternative methods were not so feasible (Mezna and Lawrence, 1994). It consists of a reaction well containing two electrodes both of which are submerged in an electrolytic solution (Figure 7.1). An amplitude stable AC signal (V_i) from a low impedance source is fed into the conductivity cell, which acts as the input impedance for an operational amplifier. Under these conditions the voltage generated at the output (V_o) is a linear measure

of the current flowing through the cell and hence is a linear measure of conductance ($1/R_i$). This can be written as $V_o = V_i * R_o/R_i$ or $V_o = \text{conductivity} * k$, where conductivity = $1/R_i$ and $k = V_i/R_o$. R_o is variable resistance from a potentiometer that is used to balance current flow. The AC signal is compared with a reference signal in a rectifying circuit, which gives the sign and magnitude of the difference, and the output voltages are converted to a 13-bit digital response. A reaction event is expressed by a change in the conductivity of the solution from a base value recorded prior to initiating the reaction.

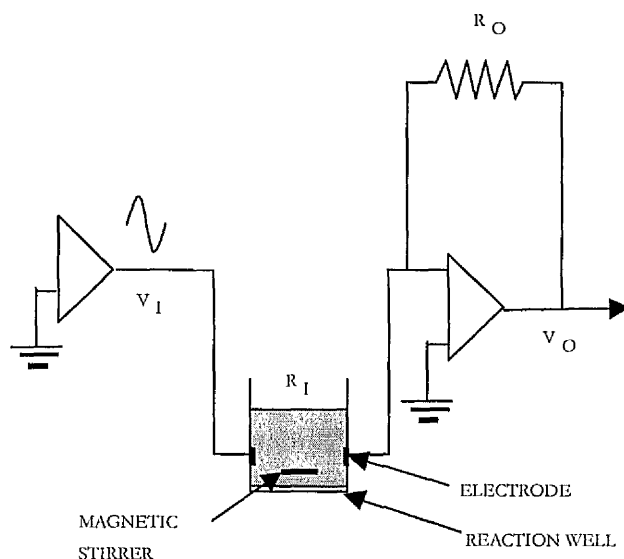


Figure 7.1 A simplistic circuit diagram for the enzymometer. Current flows into the reaction well from an AC source (V_i). The flow of current at the electrode in both directions prevents electrode polarisation. R_o is resistance from a potentiometer that balances current flow to the output voltage (V_o) as a result of changing conductance ($1/R_i$). A magnetic stirrer minimises the contribution of diffusion to conductance change in the reaction well. The temperature of the solution is kept constant by placing the well in a large volume of water that has its temperature controlled by a thermistor.

The enzymometer possesses eight small reaction wells (Figure 7.2) immersed in a heated water bath that is continuously stirred to maintain a constant temperature. The stirring motion is provided by a rotating magnetic wheel, which is also used to spin a magnetic stirring pellet

present at the bottom of each reaction well. This ensures that the reaction proceeds immediately.

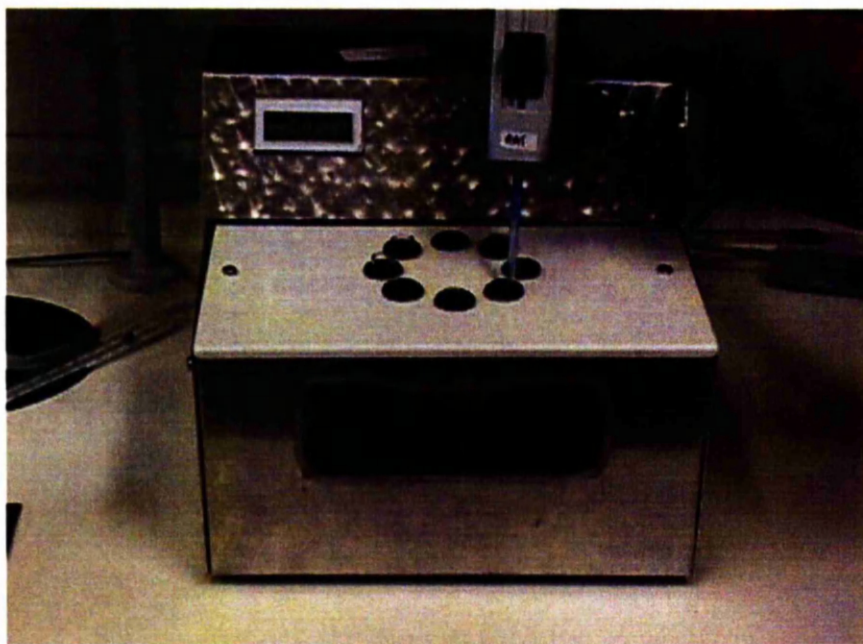


Figure 7.2 A frontal view of the enzymometer showing eight glass reaction chambers with substrate being pipetted into one of them. The unit is filled with water, which is continuously stirred and maintained at a fixed temperature.

This chapter outlines a protocol that was developed for using the enzymometer to monitor toxin activity by measuring the changes in conductance of the solution as a result of toxin interaction with the erythrocyte membrane. It is based on a similar protocol that was used to monitor the haemolytic activity of lysolecithin and linoleic acid (Lawrence *et al.*, 1974). Erythrocytes are suspended in a low conductance medium (e.g. 10 mM MOPS⁻/Na⁺ and isotonic sucrose). As ion leakage occurs, there will be a rise in conductance of the medium until the ions are equilibrated across the membrane. Leakage that occurs naturally in the absence of lytic agents is known as basal leakage. If a lytic agent is introduced, leakage will be greatly accelerated, reflected by the rate of change of conductance which is used as a measure of the potency of the lytic agent. The experiments presented in this chapter show:

1. That an increase in conductance at the initial stage of a reaction between toxin and erythrocytes corresponds with a leakage event.

2. The rise in basal leakage of ions from human and horse erythrocytes induced by pneumolysin is greater than from rabbit and sheep erythrocytes.
3. That toxin induced haemolysis of rabbit erythrocytes is preceded by a leakage event.
4. The regulation of toxin activity by both Ca^{2+} and Mg^{2+} ions.
5. That there is a drop in conductance (below basal leakage) at low toxin concentrations.

7.2 MATERIALS AND METHODS

Horse and sheep blood was purchased from E & O laboratories, Scotland, while rabbit and human were kindly provided by Diagnostics Scotland and the Glasgow Blood Transfusion unit respectively. All other chemicals were obtained from Sigma, UK. Except for human, all other blood comprised a 50% (v/v) mixture of modified Alsever's solution containing the antibiotics chloramphenicol (0.349 g/L) and neomycin sulphate (0.1 g/L). Human blood was centrifuged at 3000 rpm for 5 minutes at 20 °C and supernatant and buffy coat were removed carefully by suction.

7.2.1 SAMPLE PREPARATION

Blood was spun at 13,000 rpm in a bench-top centrifuge for one minute. The red cell pellet was washed three times in isotonic saline (154 mM NaCl) to ensure as many plasma components and preservatives as possible were removed. The pellet was resuspended in isotonic saline to yield a 30% (v/v) cell suspension, which was used as working stock. Pneumolysin was obtained from the hydrophobic interaction chromatography purification described in Chapter 4. The toxin was diluted in 10 mM MOPS (3-(*N*-morpholino) propanesulphonic acid) pH 7.2 to give a final concentration of approximately 10 μM . Digitonin, a steroid glycoside derived from seeds of *Digitalis purpurea*, permeabilises the plasma membrane of eukaryotic cells by binding to cholesterol and other unconjugated β -hydroxysterols (Elias *et al.*, 1978). It was used in the subsequent set of experiments to completely lyse red cells. Digitonin was prepared at 2 mg/ml by vortexing with isotonic sucrose in 10 mM MOPS at room temperature. Although it is poorly soluble in aqueous solutions (Mooney, 1988), 5 μl was found to be sufficient to completely lyse 2 ml of a 0.3 % (v/v) suspension of erythrocytes.

7.2.2 CONDUCTANCE MEASUREMENTS

Isotonic sucrose (0.3 M) buffered with 10 mM MOPS was degassed by heating the solution and then exposing it to a vacuum until vigorous bubbling occurred. 2 ml of the degassed buffer was added to all of the wells followed by 20 μ l of erythrocyte cell suspension to make a final cell suspension of 0.3 % (v/v). This was left to incubate for approximately 30 s so that all the wells could be blanked (i.e. gave the same reading) before addition of other components such as salts and toxin, the quantities of which can be found in the appropriate results sub-sections. Readings (in arbitrary units - Conductance) were taken prior to the addition of the last sample (this was usually pneumolysin) to monitor the start of the reaction. The temperature control was always set at 37 °C. The cells were washed several times between each successive measurement, initially with a water/ detergent mix and finally with distilled water to remove residual detergent²⁷.

7.2.3 DATA ANALYSIS

Data were displayed on software written specifically for this instrument, which showed the change in conductance²⁸ with time. Measurements were taken every 8 s but the first 24 s of the reaction (this was longer in some reactions) was not included in the analysis because of noise caused by the introduction of new solution into the reaction well. A reference well containing only cell suspension was used to measure basal leakage for those experiments investigating toxin induced leakage. Basal leakage was subtracted from conductance measurements taken for toxin induced leakage. The data were analysed using the graphing package *PRO-FIT*, (Cherwell Scientific publishing). The maximum rate of lysis was estimated using those data points at which the gradient was steepest. Using the Levenberg-Marquadt least squares algorithm, a straight line was fitted through those data points.

7.2.4 PRECAUTIONARY MEASURES

As stated earlier, leakage of ions from erythrocytes results in an increase in conductance. However, non-specific reactions can also contribute to increasing conductance. One example is the reaction of atmospheric CO₂ with MOPS⁻ that results in the formation of MOPS(H) and HCO₃⁻, the difference being responsible for the increase in conductance.

²⁷ An additional methanol wash was used for reaction cells that had digitonin added to them

²⁸ Conductance measurements were recorded in arbitrary units – conductance units.

Measurements were also sensitive to changes in temperature and volume, therefore the following precautions were taken:

- Volume additions were made using a Hamilton syringe (Bonaduz, Switzerland) fitted with a ratchet mechanism that would move the plunger by increments of 0.5 μl or 1.0 μl for precision volume additions.
- The tip of the syringe needle was submerged into the solution contained inside the reaction well to ensure that all the reactant from the syringe had completely mixed with the solution.
- Readings from each cell were taken once every 8 s. To minimise effects on conductance measurements caused by the addition of reactant, the reactant was injected immediately after a reading was taken giving time for the solution to restabilise.
- To minimise the change in temperature, the volume added to the 2 ml cell suspension was less than fifteen μl .

By following these precautions, any fluctuation in conductance was likely to be attributable only to the chemical properties of the solution. To check the significance of the chemical properties, individual chemical components used regularly in the experiments described below were tested for their change in conductance.

7.3 RESULTS

7.3.1 CONDUCTANCE PROPERTIES OF CONSTITUENT ELEMENTS

The prime interest for this study was to identify those chemicals (used in the set of experiments described in this chapter) and their combinations that would contribute to a significant change in conductance and to determine the suitability of horse and human erythrocytes for conductimetric experiments. Seven different conditions were tested: (i) buffered sucrose only; (ii) buffer + pneumolysin; (iii) buffer + cation (0.5 M CaCl_2); (iv) buffer + sheep blood (30% (v/v) erythrocytes); (v) buffer + rabbit blood (30% (v/v)); (vi) buffer + blood (30% (v/v) from the specified mammal) + cation; (vii) buffer + rabbit blood (30% (v/v)) + toxin. The chemical components for each condition were added to all eight wells so that the variation in leakage between cells with identical components would be known. The quantity of each component added for the seven different conditions is given in Table 7.1. Measurements were taken for approximately ten minutes.

Chemicals	Conditions						
	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
Sucrose (ml)	2	2	2	2	2	2	2
30% v/v erythrocytes (μ l)				20	20	20	20
0.5 M CaCl_2 (μ l)			2			2	
10 μ M pneumolysin (μ l)		2					2

Table 7.1 List of chemicals and the volumes used to test the conductance properties for seven different solution compositions. Components were added in row order (top to bottom). Erythrocytes from human, horse, rabbit and sheep were used in condition (iv), (v) and (vi), while condition (vii) used only rabbit erythrocytes.

Except for reaction wells containing blood, the change in conductance for those constituent combinations listed in Table 7.1 was found to be negligible as shown by Figure 7.3A. For reactions containing blood without toxin, both human and horse blood caused significantly higher changes in conductance in comparison with rabbit and sheep. The addition of calcium to rabbit erythrocytes at a concentration of 0.5 mM reduces ion leakage slightly (Figure 7.3A condition (vi)). Although this reduction was insignificant when compared with the change in conductance caused by the addition of pneumolysin (Figure 7.3B condition (vii)), its occurrence has also been witnessed previously by Bashford *et al* (1988). They showed that certain divalent ions were effective at blocking physiological leakage from erythrocytes, with calcium and zinc ions being more effective than magnesium. The mechanism for this blockage is thought to involve the formation of small channels caused by negatively charged membrane components repelling each other. So the binding of divalent ions to these components is thought to minimise repulsion.

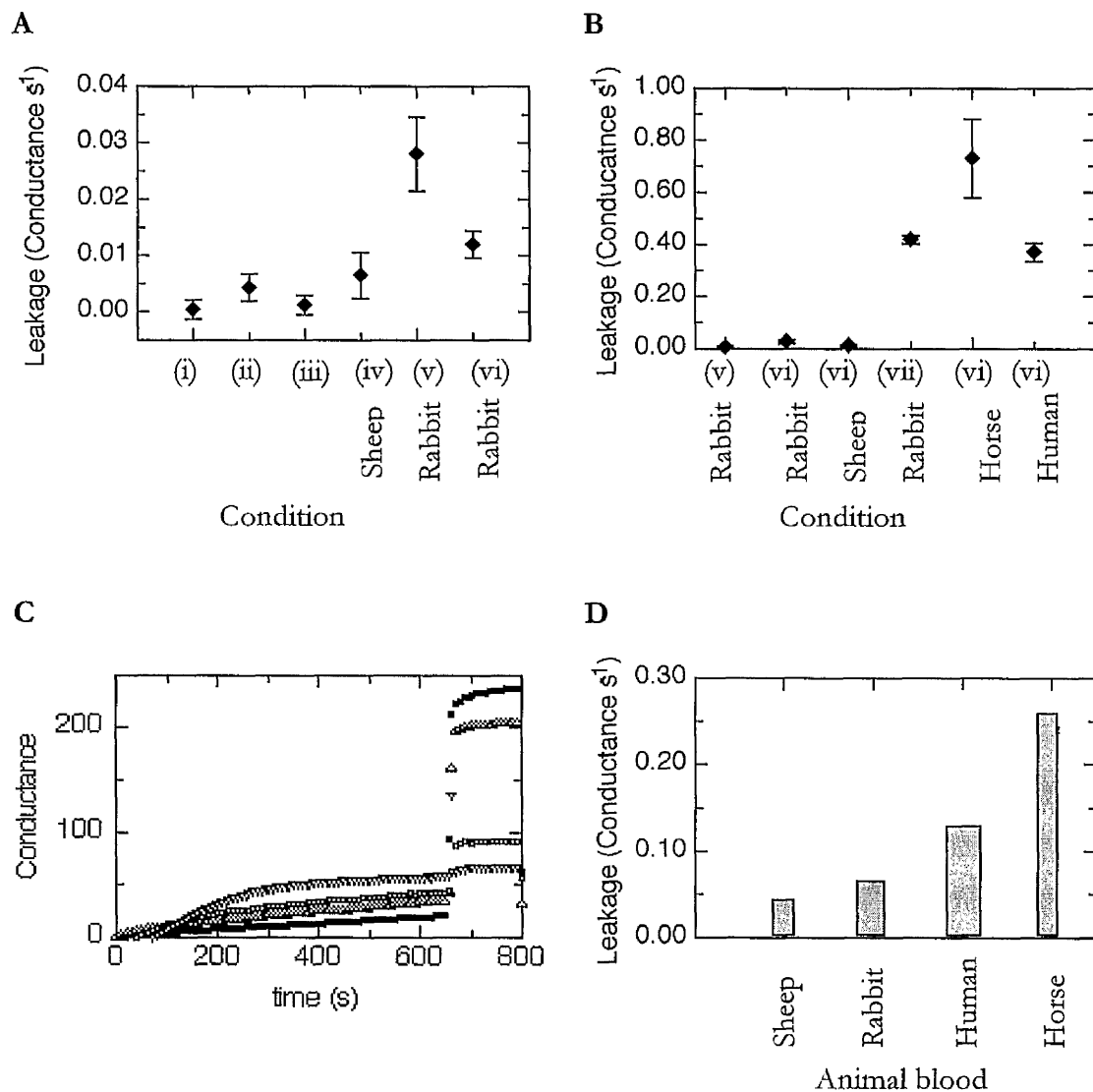


Figure 7.3 Leakage rates were measured as the change in conductance over a period of time. Leakage rates for conditions (i) to (vii) were plotted on two separate graphs (A) and (B) to illustrate how certain compositions can affect leakage rates. Those conditions including blood have been labelled below the abscissa axis with the animal name. (C) Condition (vi) was repeated for sheep (■), rabbit (▲), human (●) and horse (▼) erythrocytes, but with the addition of digitonin after 650 s. Rate of leakage from this repeat experiment is shown in (D).

A direct comparison of the leakage between the four different blood types was examined by repeating condition (vi) (Figure 7.3C), showing horse to have the highest leakage rate and sheep to have the lowest (Figure 7.3D). Because leakage from red cells is also occurring

during storage, the leakage rates may decrease slightly with successive experiments. This is seen particularly with horse and human blood, which in the first measurements shown in Figure 7.3B has a leakage rate of 0.74 and 0.38 Conductance s^{-1} respectively, but then drops to 0.24 and 0.12 Conductance s^{-1} in Figure 7.3D. No significant change was found for either rabbit or sheep blood, suggesting that it was much more stable under buffered conditions. When 5 μ l of digitonin was added to each of the four wells containing different types of blood (Figure 7.3C), both sheep and rabbit solutions showed a significantly higher rise in conductance than human and horse blood. The drop in leakage rates of horse and human blood after a short period of time indicated that both types of blood had a high degree of ion leakage when stored in isotonic saline. The minor increase in conductance upon addition of digitonin to both horse and human red cells indicated that the movement of ions across the membrane had approached (for horse erythrocytes) or was approaching (human) a steady state with the bulk solution. The findings indicate that in comparison to rabbit and sheep erythrocytes, human and horse erythrocytes have a higher ion leakage, which does not result in any apparent lysis of the red cells.

Comparative NMR measurements on the permeability to water of certain mammalian erythrocyte membranes have shown that human erythrocytes have a slightly higher water permeability than sheep erythrocytes but a much lower permeability than rabbit erythrocytes (Benga and Borza, 1995). The lack of correlation between water permeability and ion leakage for rabbit and human erythrocytes may be attributable to specific water transporters - aquaporins: a group of small proteins (28-31 kDa) that form channels in mammalian cells, specifically for the passage of water (Agre *et al.*, 1993). It may well be that human and sheep erythrocytes possess fewer water transporters in their membrane than rabbit erythrocytes. A method involving the use of PCMBs (*p*-chloromercuribenzenesulfonate) for investigating the variability of transporters between species was suggested by Benga and Borza (1995). PCMBs binds to the aquaporin CHIP28 (channel forming integral membrane protein of 28 kDa), inhibiting its ability to act as a water transporter. The degree of inhibition could be used to quantify the number of CHIP28 molecules on red blood cells of that species. Their results showed that there was a general trend in erythrocyte permeability, with high permeability cells (e.g. from rabbit) being more sensitive to PCMBs than lower permeability cells (e.g. sheep and human). However there were exceptions which implied that the correlation between permeability and number of transporters need not necessarily exclude other factors that also contribute to membrane permeability, such as the interaction of membrane proteins with lipids; the lipid composition and asymmetry of the bilayers. The table below contains data for the lipid distribution of various mammalian erythrocytes (Nelson, 1967), including human (Farquhar, 1962).

PNEUMOLYSIN-INDUCED HAEMOLYSIS

A

Species	Total lipid (g/cell) X 10 ⁻¹³	Percent of total lipid			
		Neutral lipid	Glycolipid	Ganglio- side	Phospho- lipid
Rabbit	4.15	28.9	0.8	4.5	65.8
Horse	2.58	24.5	8.0	15.5	52.0
Sheep	1.62	26.5	2.5	7.8	63.2
Human	*	-	-	-	59.0

B

Phospholipids	Rabbit	Horse	Sheep	Human
Phosphatidic acid	1.6	< 0.3	< 0.3	
Phosphatidyl ethanolamine (PE)	31.9	24.3	26.2	29.0
Phosphatidyl serine (PS)	12.2	18.0	14.1	10.0
Phosphatidyl inositol (PI)	1.6	< 0.3	2.9	-
Phosphatidyl choline (PC)	33.9	42.4	-	36.0
Sphingomyelin (SM)	19.0	13.5	51.0	21.0
Lysophosphatidyl choline (LPC)	< 0.3	1.7	-	-
Unidentified	-	-	-	3.0

Table 7.2 The data in tables A and B were obtained from Nelson (1967) (Nelson, 1967) for non-human erythrocytes and Farquhar (1962) for human erythrocytes. Table A shows a percentage distribution of the different types of lipids while table B contains the percentage composition of phospholipids. Some of the values obtained for rabbit red blood cells appear to be quite different from those obtained from Kazennov (1998) (Kazennov *et al.*, 1998). Their results indicate PC = 47.0 ± 2.4 %; PE = 15.8 ± 1.3 %; PS = 12.9 ± 1.0 %; PI = 2.8 ± 0.4 %; SM = 18.9 ± 1.7 %; LPC = 2.6 ± 0.6 %. Total lipid extract for human was 510 ± 51 mg/100 ml cells. This value was not included in table A because it was measured differently (*).

The relationship between lipid composition and water leakage was shown by Benga and Borza (1995) to relate to the PC and SM content of the membrane. A higher PC and lower SM content corresponded to a higher leakage. Similar studies on anion leakage using phosphate

ions have shown that phosphate uptake for human is 113 ± 12 nmoles/g erythrocytes per minute, which is considerably higher than for horse and sheep, which take up 24 ± 6 and 50 ± 4 nmoles/g erythrocytes per minute respectively (Deuticke and Gruber, 1970). These studies also showed a correlation between anion uptake and membrane phospholipids, in particular PC and SM. The flux appears to increase with membranes having high amounts of PC and low SM. Results from water leakage and anion uptake both show similar patterns with PC and SM content, and this correlation can be extended to the behaviour between horse and sheep erythrocytes (at least in the case of SM). But the results for rabbit and human erythrocytes do not obey this pattern. Less convincing were those phospholipids found on the inner side of the membrane (PE and PS), that appeared to show no correlation with the leakage results. Kazennov *et al.* (1998) had shown that the ion transport protein, Ca^{2+} -ATPase present in animal cells such as erythrocytes, positively correlated with the phospholipid content of the inner lipid bilayer (Kazennov *et al.*, 1998). This was not found to be the case with another ion transport protein, Na^+ , K^+ -ATPase. The activities of both proteins were found to vary between different species and were also regulated by a variety of different factors. For instance, the Ca^{2+} -ATPase is regulated by calmodulin where as the Na^+ , K^+ -ATPase is controlled by intracellular phosphates and divalent cations; membrane proteins and other calcium and anion transporting systems; chemicals that act extracellularly, such as ouabain²⁹. Although these factors were taken into account in their results, it and the previous examples illustrate the complexity of attempting to correlate the different factors responsible for ion leakage.

Although no correlation could be found between ion leakage and the factors that regulate it, the results presented in this section have shown that there is a negligible change in conductance for substrates not containing blood. A significant rise in conductance was observed in human and horse erythrocytes; and with rabbit erythrocytes but in the presence of pneumolysin. The low basal leakage in both rabbit and sheep erythrocytes and their sensitivity to pneumolysin has made them suitable candidates for subsequent conductance studies.

7.3.2 CHARACTERISTICS OF TOXIN INDUCED LYSIS IN RABBIT ERYTHROCYTES

There are six major protein groups in blood serum: fibrinogen, albumin, α_1 , α_2 , β , and the γ globulins. All of these groups possess varying degrees of negative charge. Haemoglobin a

²⁹ Ouabain is a steroid glycoside from the seeds of *Strophanthus gratus*. It inhibits sodium potassium exchange by dephosphorylating the Na^+ , K^+ -ATPase pump.

major protein constituent in mammalian erythrocytes and is present in several isoforms depending upon the animal and its stage of development. Human haemoglobin contains approximately 574 residues; is about 55 Å in diameter with a molecular weight of 64 kDa. Toxin induced leakage is thought to be caused by the formation of pores which then result in the release of molecules such as inorganic ions and proteins from inside the cell. Electron microscopy images of the pneumolysin pore indicate that the inner diameter of the pore is between 23 to 38 nm and that a 40 kDa dextran polysaccharide can offer some protection against haemolysis (Korchev *et al.*, 1998). Based on this, it is conceivable that haemoglobin could pass through toxin induced pores without disrupting the complete integrity of the cell. If this is the case then haemoglobin and other charged molecules would also contribute to conductance readings, but because of their size, the contribution would be minimal.

Studies conducted on tetanolysin (from *Clostridium tetani*) by Blumenthal and Habig (1984) and streptolysin O (from group A streptococci) by Duncan (1974) have shown that leakage of haemoglobin and ions occurs almost simultaneously. Pore forming toxins like pneumolysin belong to the same family of thiol activated cytolysins and therefore these findings have direct relevance to the results described in this section. The addition of toxin to rabbit erythrocytes in the previous experiment resulted in a significant increase in conductance. This study aims to provide some insight into whether ion leakage induced by the toxin is distinct from haemoglobin leakage.

Measurements taken from the previous experiment showed that an increase in conductance from erythrocytes in the absence of any lytic agent was attributed to the leakage of inorganic ions – basal leakage. To account for this leakage, subsequent experiments were carried out on a maximum of seven wells so that the control well only contained cell suspension (see section 7.2.3). 2 µl of pneumolysin was added to seven of the wells containing cell suspension. A 1 ml sample was taken from each well at specific periods during the reaction and kept on ice. All the samples were spun at 13,000 rpm on a bench top centrifuge and the absorbance at a wavelength of 541 nm (A_{541}) was taken for the supernatant to determine the amount of haemoglobin present in solution. 5 µl of digitonin was added to the control sample at the end of the experiment to show that the A_{541} for complete cell lysis was 0.64 absorbance units and that maximum conductance was 228 conductance units. A separate reaction to determine the total amount of haemolysis from 10 nM pneumolysin showed that after 600 s, the A_{541} of the supernatant was 0.42, indicating that at 10 nM pneumolysin, only 65% of the total haemoglobin content had been released into solution.

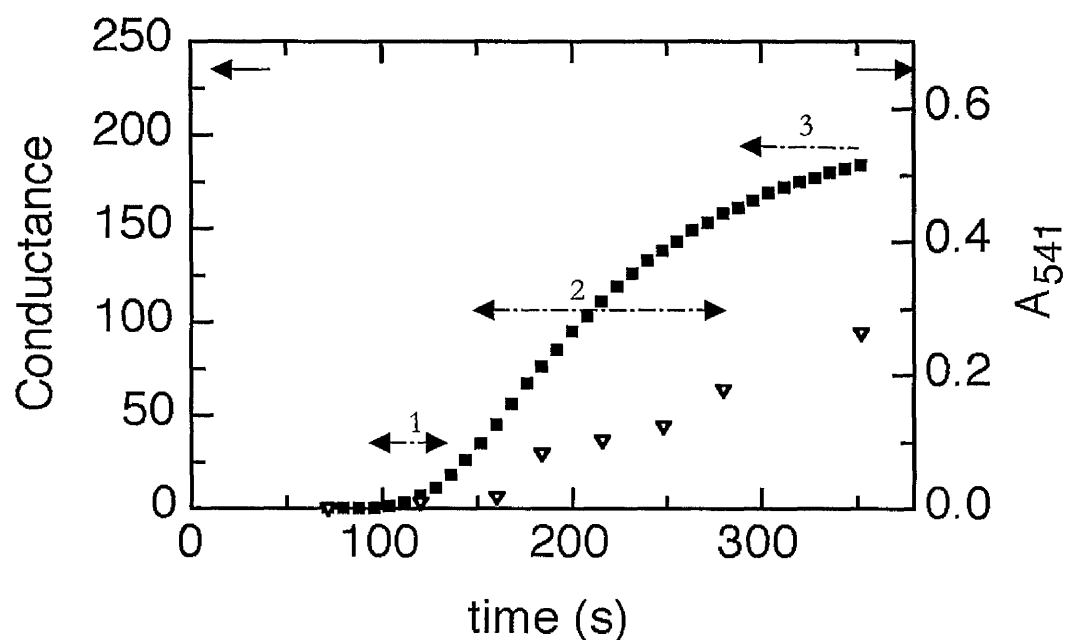


Figure 7.4 Leakage and lytic events in rabbit erythrocytes monitored by conductimetry (■) and absorbance (▼) measurements at 541 nm. Complete lysis occurred at $A_{541} = 0.64$ and Conductance = 228, indicated by the solid arrows. Change in conductance was categorised as three separate phases as indicated by the dashed arrows: (1) lag; (2) exponential and (3) stationary.

The measurements shown in Figure 7.4 show the three different phases of leakage that probably correspond with the following events. Phase 1, the lag phase is the period during which toxin remains free in solution prior to binding to membranes of the red blood cells. In phase 2, the exponential phase, a sharp rise in conductance corresponds with toxin pore formation in the membrane. During phase 3, the stationary phase a minimal change in conductance signifies that leakage has reached completion. The detection of haemoglobin during the exponential phase could either be due to the leakage of haemoglobin from pores in the cell membrane or due to its release upon cell lysis. The two events could be differentiated by performing light scattering experiments (see Harris *et al.* (1991) and Gilbert (1998)), which would indicate if the cells were intact and by measuring the absorbance spectra at 541 nm to indicate whether haemoglobin had indeed leaked from the cells. If the detection of haemoglobin outside the cell (from absorbance measurements) corresponds with cell disruption (from scattering experiments), then it could be assumed that haemoglobin leakage is consequential upon cell lysis.

By 380 s, solution conductance was approaching steady state while haemoglobin release was steadily increasing. Although the release of haemoglobin contributes to the conductance measurement, its effects are insignificant compared with that of inorganic ions due its much larger size and relatively low abundance. This is evident from the very slight increase in conductance during the stationary phase (beyond 380 s; data not shown), while haemoglobin release from the erythrocytes continues to rise.

The 30 s lag represents the shortest time for haemoglobin release into bulk solution. It is likely that the lag phase is longer because the reaction may have continued even after quenching with ice and during the centrifugation of the sample. Therefore the haemoglobin leakage curve should theoretically have a lower absorbance value than has been plotted. The leakage-lysis graphs from Blumenthal and Habig (1984) and Duncan (1974) show that ion leakage does marginally precede haemoglobin release. However because the difference was only marginal, their findings led them to conclude that a colloid osmotic mechanism³⁰ was unlikely to be the cause of lysis. However this conclusion is contrary to the results presented in this section and to the forward angle light scattering studies performed by Harris *et al.* (1991) which showed that erythrocytes in the presence of theta toxin (perfringolysin O) from *Clostridium perfringens* underwent a volume expansion prior to lysis. This finding was supported by another experiment from the same study where the erythrocyte-toxin mixture placed in a more hypertonic medium (using sucrose) offered the cell slightly more protection against lysis adding further support to an osmotic imbalance being responsible for lysis.

7.3.3 THE EFFECT OF CALCIUM AND MAGNESIUM ON TOXIN INDUCED LEAKAGE

Calcium is prevalent in most natural solvent environments. Its concentration in fresh water and within the cytoplasm of cells is approximately 0.1 μM whereas in the bloodstream it can exceed 1 mM (Williams, 1974). Magnesium ions are present in and around cells at a concentration of about 1 mM. Magnesium is prevalent in blood and is required for translocase activity at 2 mM (Bibitol *et al.*, 1987). Virtually all animal cell lipid membranes have a much higher permeability to water than to ions. For example, Ehrlich ascites tumor cells have a water permeability that is five orders of magnitude higher than for ions (Hoffmann and Mills, 1999). The rapid rate of exchange means that the movement of water across the membrane will reach a steady state with the bulk solution far sooner than the ions. So once equilibrium has been established, factors that effect movement of water include an

³⁰ This mechanism involves an initial permeability change to ions and small molecules across the cell membrane, followed by a loss in osmotic balance with macromolecules inside the cell, causing the cell to swell and eventually lyse.

imbalance of ions across the membrane. Under normal conditions, the osmolarity is maintained by a variety of factors (see section 7.3.1), which includes the Na^+ , K^+ -ATPases. If the imbalance is extreme, then changes to the cell volume may occur.

The effect of calcium and magnesium ions on pneumolysin induced leakage was investigated in two separate experiments. CaCl_2 in a volume no greater than $10\ \mu\text{l}$ was added to six of the eight wells all of which contained 0.3 % erythrocyte cell suspension. These wells contained concentrations of CaCl_2 ranging from 0.05 mM to 0.6 mM for rabbit; and 2.0 mM for sheep. All except one well (the control well, without added CaCl_2) had pneumolysin added to make a final concentration of 10 nM. The reaction was initiated by the addition of pneumolysin but measurements were taken prior to addition. A similar procedure was used to investigate the effects of toxin induced leakage at magnesium ion concentrations ranging from 0.025 to 1 mM MgCl_2 .

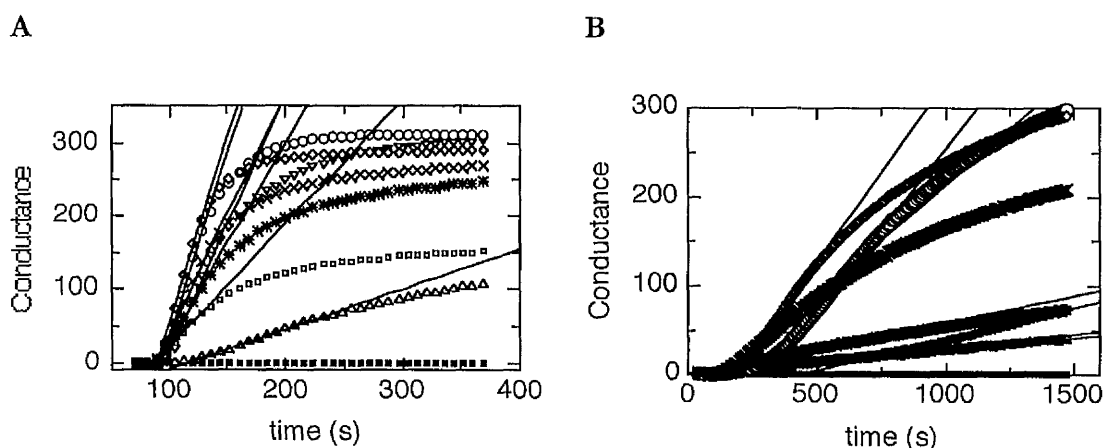


Figure 7.5 Leakage response in both rabbit and sheep erythrocytes as a result of varying CaCl_2 concentration. **A:** 0.3% v/v rabbit erythrocytes with $1\ \mu\text{l}$ pneumolysin (Δ) and with CaCl_2 at concentrations 0.05 mM (\circ), 0.1 mM (∇), 0.25 mM (\diamond), 0.4 mM (\square), 0.5 mM (\times) and 0.6 mM (\ast). **B:** 0.3% v/v sheep erythrocytes with $2\ \mu\text{l}$ pneumolysin (Δ), and then with CaCl_2 concentrations at 0.05 mM (\circ), 0.1 mM (∇), 0.2 mM (\diamond), 0.5 mM (\times), 1.0 mM (\ast) and 2 mM (\ast).

Figure 7.5 shows the rise in conductance in response to addition of toxin. Except for the well containing 0.05 mM CaCl_2 and no calcium, the conductance readings for the other five wells containing toxin appeared to have stabilised (or approached stability) at just over 300 conductance units. At 0.05 mM CaCl_2 , conductance rose fairly rapidly between 100 – 150 s

and then settled to a slower linear rise, which corresponded to basal leakage. The reaction with no added CaCl_2 showed an even slower rate of increase, settling at the basal leakage rate a short period after the 0.05 mM CaCl_2 reaction curve had settled. In the range of concentrations tested, optimum haemolytic activity was observed at calcium concentrations in the region of 0.3 mM calcium for rabbit erythrocytes and around 0.5 mM for sheep erythrocytes (Figure 7.6). Leakage was inhibited at higher calcium concentrations.

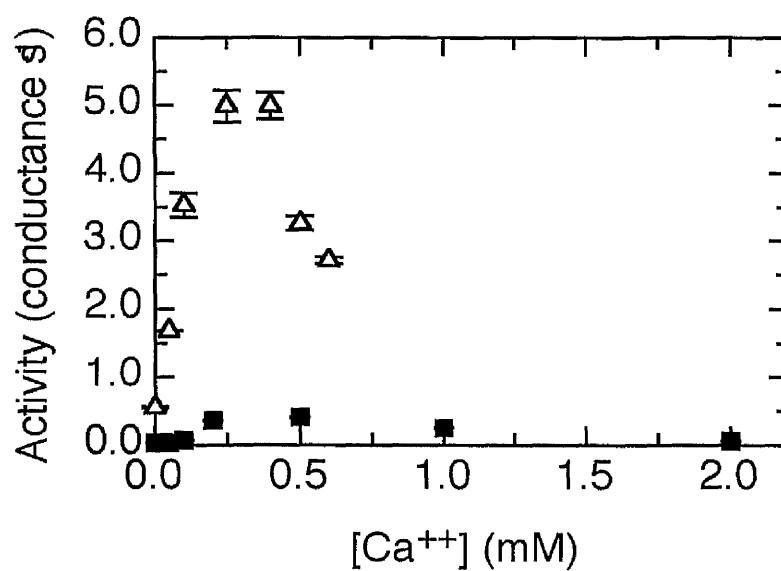


Figure 7.6 Leakage rates taken from the gradients in Figure 7.5 were plotted against CaCl_2 concentrations for both rabbit (Δ) and sheep (\blacksquare) erythrocytes. If leakage assumes a Gaussian distribution, optimum leakage should occur at a calcium concentration between 0.25 to 0.4 mM CaCl_2 in rabbit erythrocytes and between 0.3 to 1.0 mM CaCl_2 in sheep erythrocytes.

These results clearly show the activity of the toxin on rabbit erythrocytes to be significantly more amplified than sheep erythrocytes at CaCl_2 concentrations less than 0.3 mM. They also show that the activation and inhibition kinetics in sheep erythrocytes are slightly different. This is made clearer in Figure 7.7, where the graphs showing activation kinetics appear to have a longer lag phase than those showing inhibitory kinetics. Assuming the lag phase corresponds with toxin insertion, and the exponential phase corresponds with pore formation, then it would appear that pore formation occurs very quickly at higher calcium concentrations. One suggestion is that the increasing calcium concentration stimulates specific transport ATPases to pump out intracellular ion stores. This and the integration of the toxin into the membrane may have a combined effect in reducing the lag phase. This can be tested by

monitoring ion leakage at the range of calcium concentrations but without the toxin. It is obvious that no clear explanation can be provided from these results but what has been shown should support the need for further investigation into the activating and inhibiting properties of the cation.

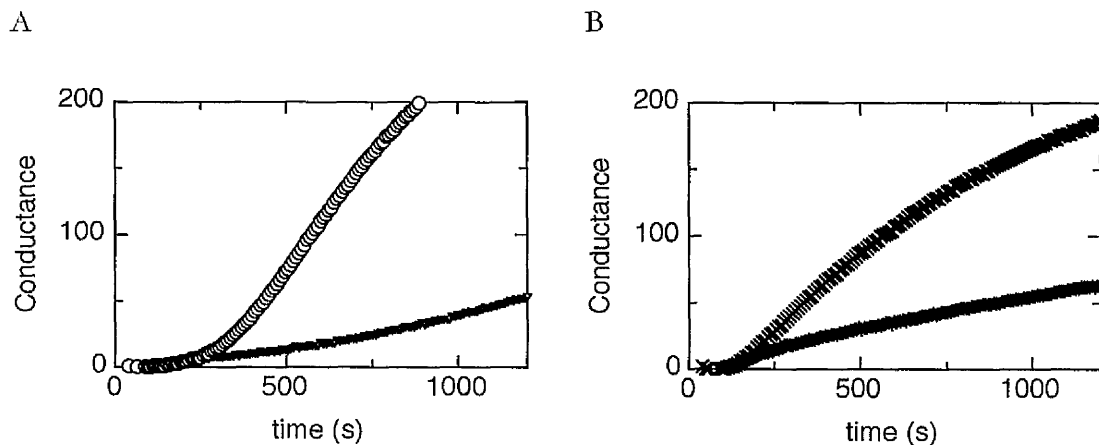


Figure 7.7 Four slopes taken from graph **B** in Figure 7.5, at 0.1 mM (∇), 0.2 mM (\circ), 1.0 mM (\times) and 2 mM ($*$) CaCl_2 were used to illustrate calcium activation (**A**) and calcium inhibition (**B**).

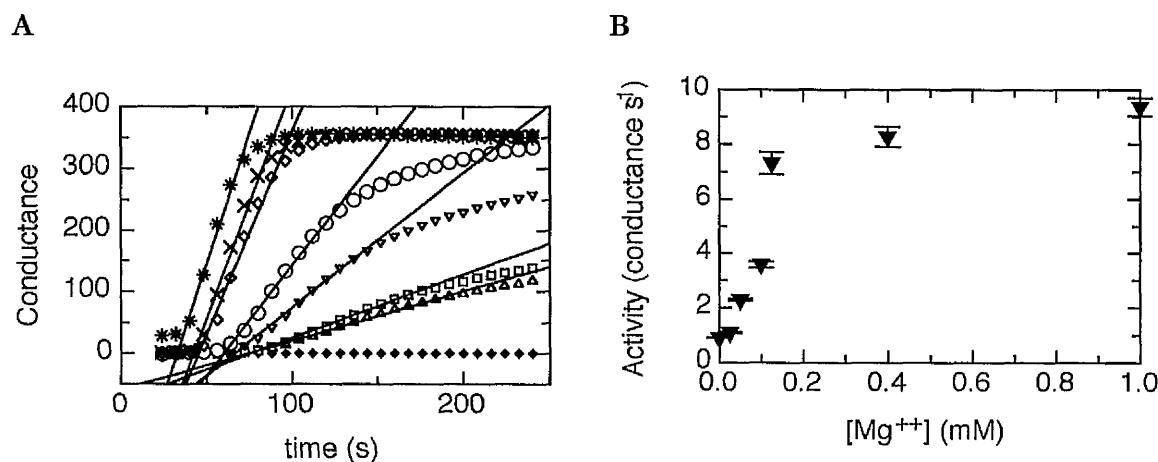


Figure 7.8 **A**: 0.3% v/v rabbit erythrocytes with 10 μl pneumolysin. Conductance of the chamber was monitored at the following $[\text{Mg}^{++}]$: 0 mM (\wedge), 0.025 mM (\square), 0.05 mM (∇), 0.1 mM (\circ), 0.125 mM (\diamond), 0.4 mM (\times) and 1 mM ($*$). **B**: Maximum haemolytic activity occurs at $[\text{Mg}^{++}]$ greater than 0.2 mM and remains stable up to 1 mM.

In contrast to calcium induced leakage, the effect of magnesium up to 1 mM was not subject to any inhibition. The tests showed that haemolytic activity (Figure 7.8) was sensitive to

magnesium concentrations of less than 0.125 mM and at concentrations exceeding 0.4 mM activity was optimal. The lack of any inhibition at concentrations up to 1 mM was in agreement with the original findings by Johnson (1972). Korchev *et al.* (1998) also showed that inhibition occurred beyond magnesium concentrations of 100 mM.

7.3.4 PNEUMOLYSIN DOSE RESPONSE

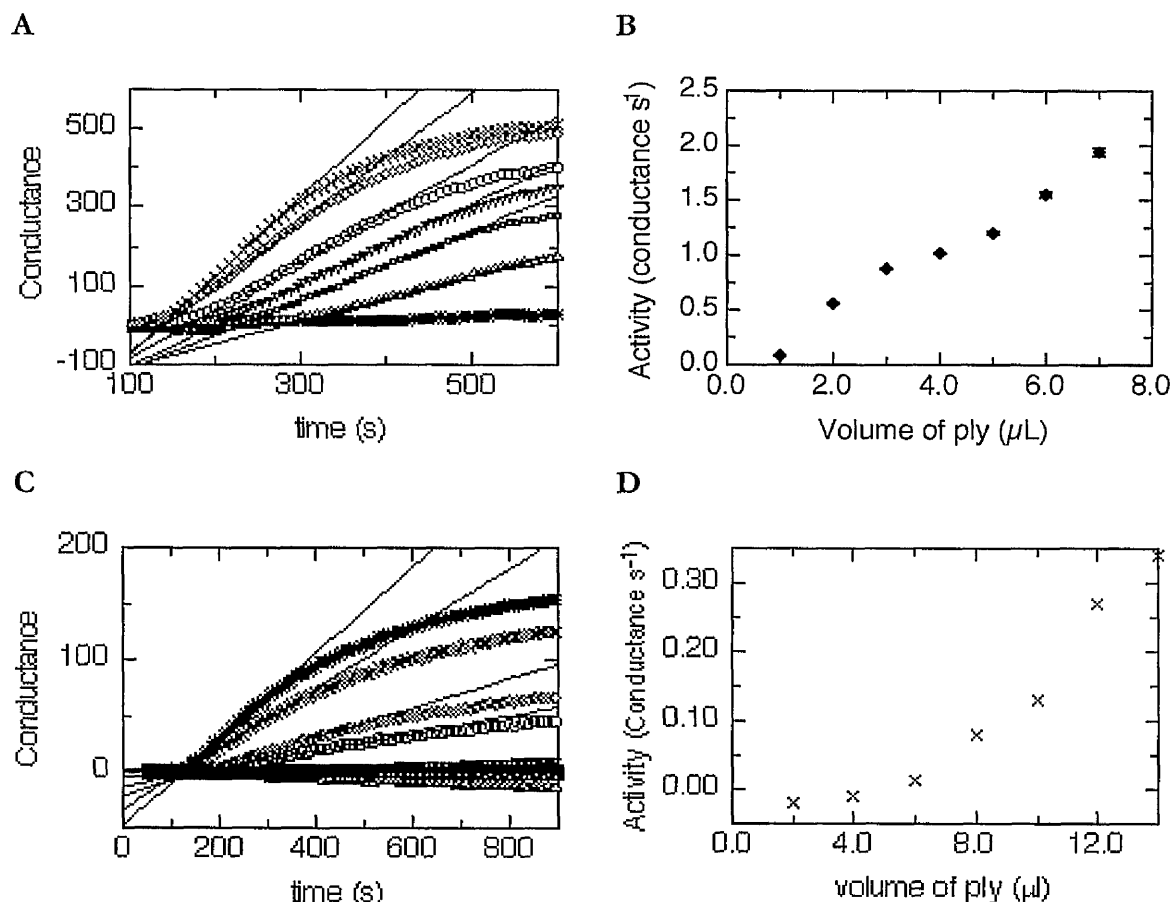


Figure 7.9 A: 0.3% v/v sheep erythrocytes containing 0.5 mM CaCl₂ (□) with 1 μL (△), 2 μL (▢), 3 μL (▽), 4 μL (◻), 5 μL (◇), 6 μL (×) and 7 μL (※) of pneumolysin. B: Increasing activity with increasing toxin concentration. C: 0.3% v/v rabbit erythrocytes containing 0.25 mM CaCl₂ (□), and 2 μL (△), 4 μL (▢), 6 μL (▽), 8 μL (◻), 10 μL (◇), 12 μL (×) and 14 μL (※) of pneumolysin at a 1:100 dilution of the stock (pneumolysin). The control had no pneumolysin (□). D: Toxin activity increases as a function of its concentration.

Numerous experiments including those described in this chapter have already shown that membrane permeability to ions is affected by the presence of lytic agents. In this study, ion

leakage is monitored at toxin concentrations ranging between 5 and 35 nM in sheep and 0.1 to 0.7 nM in rabbit. Sheep erythrocyte cell suspensions containing 0.5 mM CaCl_2 were incubated with 10 mM pneumolysin at volumes ranging between 1 to 7 μl in seven of the wells. The control well contained cell suspension with 0.5 mM CaCl_2 . The same procedure was used for rabbit cells except 0.25 mM CaCl_2 was used instead of 0.5 mM, and the toxin concentration was diluted to 0.1 mM, from which a range of volumes (between 2 to 14 μl) were added.

The effect of increasing toxin concentration with sheep cells was a linear increase in conductance showing first order kinetics (Figure 7.9A and B). The same was found with rabbit cells containing toxin at volumes greater than 6 μl . At lower doses, the conductance fell below basal levels. This drop in conductance could imply that ion leakage was partially inhibited or that ions from the bulk solution were being drawn into the cell. A variety of factors could be responsible for either effect. For reduced leakage, it may be that the toxin interacts differently with the membrane at low concentrations. This has been demonstrated to occur with the staphylococcal α -toxin, which undergoes two distinct interactions with rabbit erythrocytes depending upon its concentration. At less than 50 nM, the toxin was thought bind exclusively to a cellular receptor, where as at concentrations greater than 200 nM, it was found to absorb non-specifically to the lipid bilayer (Hildebrand *et al.*, 1991).

The possibility that the toxin can interact with the membrane so as to draw ions into the cell seems less likely. However membrane permeability can be controlled to a certain extent artificially with the addition of cholesterol or combinations of sulphhydryl reagents³¹ (Benga *et al.*, 1991). Cholesterol for instance forms strong interactions with SM possibly as a result of increased opportunities for Van der Waals interactions between the molecules. The strength of these interactions packs SM molecules closer together, tightening the membrane and thereby reducing water leakage (Lundz-Katz *et al.*, 1988). It was shown that pneumolysin binds to cholesterol within the cell membrane where it oligomerises to form a pore (Morgan *et al.*, 1995). This binding would not only disrupt the packing arrangement in the membrane but if like DTNB it also prevented anion exchange, then it could also explain the entry of ions into the cell. The evidence to support such an explanation is very weak and therefore requires further work.

³¹ Benga *et al.* (1991) have used NMR studies to show that SH-reagents such as NEM together with DTNB have an inhibitory effect on water permeability. NEM is known to inhibit glucose permeability and translocase activity while DTNB inhibits anion exchange.

7.4 DISCUSSION

Haemolysis is an ill defined process, which is usually measured by the proportion of haemoglobin released into the medium. Release could be due to all-or-nothing bursting of the cells where the reaction curve reflects the distribution of membrane fragility. In many cases the cells will reseal after bursting to form apparently intact ghosts. The other possibility is that the leakage of haemoglobin is progressive for a single cell and occurs at an increasing rate as the induced pores expand in diameter. This type of behaviour is typical of the response to fatty acids where there is progressive weakening of the membrane leading to osmotic lysis (Lawrence *et al.*, 1974). Whichever of these processes occurs, in almost all cases a substantial amount of the cell electrolyte is released into the medium before any haemoglobin is released. This means that the conductance assay measures early events in lytic damage compared with either lysis or haemoglobin leakage. The result presented in section 7.3.2 clearly shows that ion leakage precedes haemoglobin leakage and cell lysis. This finding along with work performed by Harris *et al.* (1991), was used to support the colloid osmotic mechanism for lysis.

A typical leakage assay has erythrocytes suspended in a low-conductance (10mM MOPS/Na⁺) isotonic sucrose medium at less than 0.5% v/v. Because the concentration of erythrocytes is so small, the total conductance change possible is about 2% (A.J. Lawrence, personal communication). For changes that are less than 5%, conductance is a linear measure of specific ion concentration. When applied to erythrocyte leakage the situation is complex due to the variety of ionic species in the cell. However it is simplified by the fact that about 85% of the conductance change can be attributed to potassium leakage, as observed with the potassium specific ionophores, such as valinomycin. It is probable that for cells stored in isotonic saline that the counter ion will be almost entirely chloride. Thus the majority of the early conductance leakage response will be a linear response to KCl leakage. In contrast the last 15-20% of the response will be due to a variety of ions, especially the counter ions of the various proteins. It is unlikely that the proteins themselves will contribute greatly to the observed responses, primarily because of their low charge to size ratio.

Various experiments have shown that erythrocytes are more stable in isotonic sucrose than isotonic saline solutions (A.J. Lawrence, personal communication). When erythrocytes from different animal sources were incubated in isotonic sucrose medium, the rise in ion conductivity indicated basal leakage. By maintaining each reaction well at a constant temperature (37°C), the basal leakage observed in different species was attributed to the physio-chemical characteristics of the cell. The degree of reproducibility varied among different species, with rabbit and sheep being the most stable. Leakage rates increased in the

order of sheep, rabbit, mouse rat, human, donkey, horse (results from mouse, rat and donkey were not presented here). At the extreme, at 37°C horse erythrocytes would equilibrate with the medium in about 10 minutes whilst this did not occur for sheep erythrocytes within 2 hours. Even human erythrocytes leak too rapidly to be a useful experimental indicator of lytic activity. A review of the literature for phospholipid composition of the animal cells used in this study indicated no obvious correlation with leakage. It therefore seems more probable that the leakiness of the membrane is related to the way it is distorted by interaction with the submembraneous cytoskeleton.

Although inhibitory activity for different ions has been shown by a number of workers, there was no mention of leakage being amplified by increasing ion concentrations less than 0.25 mM Ca^{++} and Mg^{++} . Whether this observation was due to the presence of pneumolysin was not shown in this experiment and would require an additional experiment to that described in section 7.3.3 but without the addition of toxin. Optimal activity for Ca^{++} was found to be between 0.2 and 0.7 mM for both rabbit and sheep erythrocytes beyond which inhibition occurs. Factors other than the actual ions that contribute to inhibition may include interaction of the toxin with membrane (or even cellular) components from lysed cells. This was given as one of the explanations for the obscure trend in toxin content during the different production and purification stages of the previous study (Chapter 6). But ion concentration was not the only factor to affect leakage in erythrocytes treated with toxin. The fall in the conductance of the medium to levels below basal leakage was found to occur at very low toxin concentrations (less than 0.3 nM). Although this was attributed to properties of the toxin (see previous section) there is clearly much further work required to clarify the cause of this observation.

The findings from this study have not only shown how toxin activity is regulated by calcium ion concentration but also how its own concentration can have a reverse effect on leakage. The use of the enzymometer in conjunction with the absorbance assay showed the difference between ion and haemoglobin leakage in rabbit erythrocytes to add support to the colloid osmotic lysis mechanism for pneumolysin. These results also exemplify some key benefits of the enzymometer over other contemporary methods for monitoring erythrocyte leakage.

CHAPTER 8

CONCLUSION

In a project that centres on protein interactions, pneumolysin represents a particularly interesting system because:

- (i) it is expressed by a bacterium, *Streptococcus pneumoniae*, that has been subject of extensive work for scientific and medical reasons;
- (ii) medical research on the toxin is currently focussed on using it in vaccine development against bacterial pathogenesis;
- (iii) no experimental atomic resolution structure data are currently available for the toxin and it is therefore a good example of how structural studies can be performed in the absence of high resolution experimental data;
- (iv) it undergoes a variety of different types of interactions that includes oligomer formation due to self-interactions; interaction with and insertion into cholesterol containing membranes to form pores; interaction with anti-pneumolysin antibodies to form complexes that in some cases are capable of neutralising toxin activity;
- (v) it can be purified relatively quickly in reasonable quantities.

In the studies described in this thesis, a computer program for generating bead models (*SOMO*) was developed and successfully tested on three particularly well characterised experimental systems: bovine pancreatic trypsin inhibitor, ribonuclease A and lysozyme. The results for the theoretical hydrodynamic predictions for these models show a very good correlation with experimental values. Good hydrodynamic predictions were also obtained for the much larger, dimeric protein, citrate synthase. The next step of this project is to test the program with small-angle scattering data before applying it to studies on pneumolysin.

While development of the bead generation algorithm is almost complete, there is still considerable work required in developing *MOLed*, the molecular editor that was being developed specifically for visualising and manipulating models. A completed version of this program would be beneficial for the bead modelling community, but it may prove

more feasible to investigate other molecular visualisation programs that can be modified to perform the same role as *MOLEd*. *RAZ*, which is an extension to the *Rasmol* program was used to visualise the majority of bead models presented in this thesis.

Experimental work on the toxin has involved the development of an alternative protein purification technique using perfusion chromatography, a modification to the protocol for chemically derivatising the toxin, and a protocol for monitoring the activity of the toxin. While purified pneumolysin can be obtained in a shorter time scale to previous methods, there is still scope for further optimisation. However, based on toxin yield studies, a greater than 50% drop in pneumolysin content was observed between harvesting and disrupting bacterial cells.

Chemical derivatisation of the toxin with TNB, based on the protocol by Wang *et al.* (1996), resulted in a considerable amount of aggregation in previous work by Gilbert (1998). The extent of this aggregation was limited by derivatising the toxin after size exclusion chromatography and by only using very dilute concentrations of the toxin ($< 1 \mu\text{M}$).

By developing a protocol in which the Enzymometer could be used for monitoring ion leakage, it has been possible to show that only rabbit and sheep red blood cells were suitable for this study. Using blood from both these organisms it was possible to show that toxin activity was regulated by calcium ions, with maximum activity occurring in the presence of approximately 0.4 mM Ca^{++} . This study also adds support to the possibility that cell lysis occurs by osmotic shock; a proposition thought unlikely based on previous studies.

In conclusion, by having a working purification protocol; a method for monitoring pneumolysin activity, and a computer program for generating solution models it is expected that more results on the interactions of pneumolysin will be forthcoming.

REFERENCES

- Abola, E., Kuhn, P., Earnest, T. and Stevens, R.C. (2000) Automation of X-ray crystallography. *Nature Structural Biology* **7**, 973-977.
- Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F. and Weng, J. (1987) Protein data bank. In: *Crystallographic Databases- Information Content Software Systems Scientific Applications*, Allen, F.H., Bergerhoff, G. and R. Sievers (Eds.) Data Commission of the International Union of Crystallography, Bonn pp 107-132.
- Agre, P., Preston, G.M., Smith, B.L., Jung, J.S., Raina, S., Moon, C., Guggino, W.B. and Nielsen, S. (1993) Aquaporin CHIP: the archetypal molecular water channel. *The American Physiological Society* F463-F475
- Alder, G.M., Bashford, C.L. and Pasternak, C.A. (1990) Action of diphtheria toxin does not depend on the induction of large, stable pores across biological membranes. *Journal of Membrane Biology* **113**, 67-74.
- Altman, R.B., Hughes, C. and Gerstein, M.B. (1995) Methods for displaying macromolecular structural uncertainty - application to the globins. *Journal of Molecular Graphics* **13**, 142-152.
- Amann, E. and Brosius, J. (1985) ATG vectors for regulated high level expression of cloned genes in *Escherichia coli*. *Gene* **40**, 183-190.
- Banachowicz, E., Gapinski, J. and Patkowski, A. (2000) Solution structure of biopolymers: A new method of constructing a bead model. *Biophysical Journal*. **78**, 70-78.
- Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* **10**, 411-421.
- Barlow, P.N., Steinkasserer, A., Norman, D.G., Kieffer, B., Wiles, A.P., Sim, R.B. and Campbell, I.D. (1993) Solution structure of a pair of complement modules by nuclear magnetic resonance. *Journal of Molecular Biology* **232**, 268-284.

Bashford, C.L., Alder, G.M., Fulford, L.G., Korchev, Y.E., Kovacs, E., Mackinnon, A., Pederzoli, C. and Pasternak, C.A. (1996) Pore formation by *S. aureus* alpha-toxin in liposomes and planar lipid bilayers - effects of nonelectrolytes. *Journal of Membrane Biology* **150**, 37-45.

Bashford, C.L., Alder, G.M., Graham, J.M., Menestrina, G. and Pasternak, C.A. (1988) Ion modulation of membrane permeability: Effect of cations on intact cells and on cells and phospholipid bilayers treated with pore-forming agents. *Journal of Membrane Biology* **103**, 79-94.

Bashford, C.L., Micklem, K.J. and Pasternak, C.A. (1985) Sequential onset of permeability changes in mouse ascites induced by Sendai virus. *Biochimica et Biophysica Acta - Biomembranes* **814**, 247-255.

Bashford, C.L., Rodrigues, L. and Pasternak, C.A. (1989) Protection of cells against membrane damage by haemolytic agents: divalent cations and protons act at the extracellular side of the plasma membrane. *Biochimica et Biophysica Acta - Biomembranes* **983**, 56-64.

Belhadjmostefa, K., Poet, R. and Milner-White, E.J. (1991) Displaying inter -main chain hydrogen -bond patterns in proteins. *Journal of Molecular Graphics* **9**, 194

Benga, G. and Borza, T. (1995) Diffusional water permeability of mammalian red blood cells. *Comparative Biochemistry and Physiology* **112B**, 653-659.

Benga, G., Popescu, O., Borza, V., Hodarnau, A., Pop, V.I. and Wrigglesworth, J. (1991) Water transport in human red cells: effects of 'non-inhibitory' sulphhydryl agents. *Biochimica et Biophysica Acta* **1061**, 309-312.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Research* **28**, 235-242.

Betts, M.J. and Sternberg, M.J.E. (1999) An analysis of conformational changes on protein-protein association: implications for predictive docking. *Protein Engineering* **12**, 271-283.

Bibitól, M., Fellmann, P., Zachowski, A. and Devaux, P.F. (1987) Ion regulation of phosphatidylserine and phosphatidylethanolamine outside-inside translocation in human erythrocytes. *Biochimica et Biophysica Acta* **904**, 268-282.

Bloomfield, V., Dalton, W.O. and van Holde, K.E. (1967a) Frictional coefficients of multi-subunit structures. I. Theory. *Biopolymers* **5**, 135-148.

Bloomfield, V., Dalton, W.O. and van Holde, K.E. (1967b) Frictional coefficients of Multisubunit structures. II. Application to Proteins and Viruses. *Biopolymers* **5**, -149

Blume, M. (1975) Introduction to the theory of elastic neutron scattering. In: Schoenborn, B.P., (Ed.) *Neutron Scattering for the Analysis of Biological Structures*, pp. I-18-I-25

Blumenthal, R. and Habig, W.H. (1984) Mechanism of tetanolysin-induced membrane damage: studies with black lipid membranes. *Journal of Bacteriology* **157**, 321-323.

Bock, J.R. and Gough, D.A. (2000) Predicting protein-protein interactions from primary structure. *Bioinformatics* **17**, 455-460.

Boulnois, G.J., Paton, J.C., Mitchell, T.J. and Andrew, P.W. (1991) Structure and function of pneumolysin, the multifunctional, thiol-activated toxin of *Streptococcus pneumoniae*. *Molecular Microbiology* **5**, 2611-2616.

Buck, M., Boyd, J., Redfield, C., MacKenzie, D.A., Jeenes, D.J., Archer, D.B. and Dobson, C.M. (1995) Structural determinants of protein dynamics: analysis of ^{15}N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. *Biochemistry* **34**, 4041-4055.

Burley, S.K. (2000) An overview of structural genomics. *Nature Structural Biology* **7**, 932-934.

Butikofer, P., Lin, Z.W., Chiu, D.T.Y., Lubin, B. and Kuypers, F.A. (1990) Transbilayer distribution and mobility of phosphatidylinositol in human red blood cells. *Journal of Biological Chemistry* **265**, 16035-16038.

Buzzell, J.G. and Tanford, C. (1956) The effect of charge and ionic strength on the viscosity of ribonuclease. *Journal of Physical Chemistry* **60**, 1204-1207.

Byron, O. (1996) Size determination of proteins: A. Hydrodynamic methods. In *Proteins Labfax*, N. C. Price (Ed.) Academic Press, San Fransisco pp 161-174.

- Byron, O. (1997) Construction of hydrodynamic bead models from high -resolution X -ray crystallographic or nuclear magnetic resonance data. *Biophysical Journal* **72**, 408-415.
- Byron, O. (2000) Hydrodynamic bead modeling of biological macromolecules. *Methods in Enzymology* **321**, 278-304.
- Cantor, C.R. and Schimmel, P.R. (1980) *Part II: Techniques for the Study of Biological Structure and Function*, San Francisco: W.H. Freeman and Company.
- Carrasco, B. and Garcia de la Torre, J. (1999a) Hydrodynamic properties of rigid particles: Comparison of different modeling and computational procedures. *Biophysical Journal* **76**, 3044-3057.
- Carrasco, B. and Garcia de la Torre, J. (1999b) Improved hydrodynamic interaction in macromolecular bead models. *Journal of Chemical Physics* **111**, 4817-4826.
- Carrasco, B., Garcia de la Torre, J. and Zipper, P. (1999b) Calculation of hydrodynamic properties of macromolecular bead models with overlapping spheres. *European Biophysics Journal* **28**, 510-515.
- Chacon, P., Diaz, J.F., Moran, F. and Andreu, J.M. (2000) Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *Journal of Molecular Biology* **299**, 1289-1302.
- Chirico, G., Beretta, S. and Baldini, G. (1999) Conformation of interacting lysozyme by polarized and depolarized light scattering. *Journal of Chemical Physics* **110**, 2297-2304.
- Cundell, D., Masure, H.R. and Tuomanen, E.I. (1995) The molecular -basis of pneumococcal infection - a hypothesis. *Clinical Infectious Diseases* **21**, S204-S212.
- de los Toyos, J.R., Mendez, F.J., Aparicio, J.F., Vazquez, F., Suarez, M.D., Fleites, A., Hardisson, C., Morgan, P.J., Andrew, P.W. and Mitchell, T.J. (1996) Functional -analysis of pneumolysin by use of monoclonal -antibodies. *Infection and Immunity* **64**, 480-484.

- Deuticke, B. and Gruber, W. (1970) Anion permeability of mammalian red blood cells: Possible relation to membrane phospholipid patterns. *Biochimica et Biophysica Acta* **211**, 369-372.
- Devaux, P.F. (1990) Static and dynamic lipid asymmetry in cell membranes. *Biochemistry* **30**, 1163-1173.
- Devaux, P.F. (1992) Protein involvement in transmembrane lipid asymmetry. *Annual Review of Biophysics and Biomolecular Structure* **21**, 417-439.
- Dossick, S.E. and Kaiser, G.E. (1999) CHIME: A metadata-based distributed software development environment. *Columbia University Computer Science: technical report* **1687**, 464-475.
- Dubin, P.L., Edwards, S.L., Mehta, M.S. and Tomalia, D. (1993) Quantitation of non-ideal behavior in protein size-exclusion chromatography. *Journal of Chromatography A* **635**, 51-60.
- Dubin, S.B., Clark, S.A. and Benedek, G.B. (1971) Measurement of the rotational diffusion coefficient of lysozyme by depolarized light scattering: configuration of lysozyme in solution. *Journal of Chemical Physics* **54**, 5158-5164.
- Dughan, L. (1994) Perfusion confusion - how does high-speed chromatography work. *Bio-Technology* **12**, 1090-1090.
- Duncan, J.L. (1974) Characteristics of streptolysin O hemolysis: Kinetics of hemoglobin and ⁸⁶rubidium release. *Infection and Immunity* **9**, 1022-1027.
- Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C. and Obradovic, Z. (2001) Intrinsically disordered proteins. *Journal of Molecular Graphics and Modelling* **19**, 26-59.
- Egelstaff, P.A. (1975) Inelastic scattering of neutrons and possible biological applications. In: Schoenborn, B.P., (Ed.) *Neutron Scattering for the Analysis of Biological Structures*, pp. I-26-I-39
- Elias, P., Goerke, J. and Friend, D. (1978) *Journal of Cell Biology* **78**, 577

Errington, N., Byron, O. and Rowe, A.J. (1999) Conformational spectra - probing protein conformational changes. *Biophysical Chemistry* **80**, 189-197.

Farquhar, J.W. (1962) Human erythrocyte phosphoglycerides. *Biochimica et Biophysica Acta* **60**, 80-89.

Fischer, D. and Eisenberg, H. (1999) Finding families for genomic ORFans. *Bioinformatics* **15**, 759-762.

Gabb, H.A., Jackson, R.M. and Sternberg, M.J.E. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *Journal of Molecular Biology* **272**, 106-120.

Gallagher, W.H. and Woodward, C.K. (1989) The concentration dependence of the diffusion coefficient for bovine pancreatic trypsin inhibitor: a dynamic light scattering study of a small protein. *Biopolymers* **28**, 2001-2024.

Garcia de la Torre, J. (1989) Hydrodynamic properties of macromolecular assemblies. In: Harding, S.E. and Rowe, A.J., (Eds.) *Dynamic Properties of Biomolecular Assemblies*, Royal Society of Chemistry, Cambridge UK pp. 3-31.

Garcia de la Torre, J. and Bloomfield, V.A. (1977) Hydrodynamic properties of macromolecular complexes. I. Translation. *Biopolymers* **16**, 1747-1763.

Garcia de la Torre, J. and Bloomfield, V.A. (1981) Hydrodynamic properties of complex, rigid, biological macromolecules - theory and applications. *Quarterly Reviews of Biophysics* **14**, 81-139.

Garcia de la Torre, J. and Carrasco, B. (1998) Intrinsic viscosity and rotational diffusion of bead models for rigid macromolecules and bioparticles. *European Biophysics Journal* **27**, 549-557.

Garcia de la Torre, J., Huertas, M.L. and Carrasco, B. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophysical Journal* **78**, 719-730.

- Garcia de la Torre, J. and Bloomfield, V.A. (1981) Hydrodynamic properties of complex rigid biological macromolecules: theory and applications. *Quarterly Review of Biophysics* **14**, 81-139.
- Garcia de la Torre, J., Carrasco, B. and Harding, S.E. (1997) SOLPRO: Theory and computer program for the prediction of SOLution PROperties of rigid macromolecules and bioparticles. *European Biophysics Journal* **25**, 361-372.
- Garcia de la Torre, J., Harding, S.E. and Carrasco, B. (1999) Calculation of NMR relaxation, covolume, and scattering-related properties of bead models using the SOLPRO computer program. *European Biophysics Journal* **28**, 119-132.
- Garcia de la Torre, J., Navarro, S., Lopez, M., Diaz, F.G. and Lopez, C. (1994) HYDRO: A computer program for the prediction of hydrodynamic properties of macromolecules. *Biophysical Journal* **67**, 530-531.
- Garcia de la Torre, J., and Rodes, V. (1983) Effects from bead size and hydrodynamic interactions on the translational and rotational coefficients of macromolecular bead models. *Journal of Chemical Physics* **79**, 2454-2460.
- Gascard, P., Tran, D., Sauvage, M., Sulpice, J.-C., Fukami, K., Takenawa, T. and Giraud, F. (1991) Asymmetric distribution of phosphoinositides and phosphatidic acid in the human erythrocyte membrane. *Biochimica et Biophysica Acta* **1069**, 27-36.
- Gerstein, M. and Chothia, C. (1996) Packing at the protein - water interface. *Proceedings of the National Academy of Sciences (USA)* **93**, 10167-10172.
- Gilbert, R.J.C. (1998) The structure, conformation and mechanism of the bacterial toxin pneumolysin from *Streptococcus pneumoniae*. University of Leicester. Doctor of Philosophy.
- Gilbert, R.J.C., Jimenez, J.L., Chen, S., Tickle, I.J., Rossjohn, J., Parker, M., Andrew, P.W. and Saibil, H.R. (1999) Two structural transitions in membrane pore formation by pneumolysin, the pore forming toxin of *Streptococcus pneumoniae*. *Cell* **97**, 647-655.
- Gilbert, R.J.C., Rossjohn, J., Parker, M.W., Tweten, R.K., Morgan, P.J., Mitchell, T.J., Errington, N., Rowe, A.J., Andrew, P.W. and Byron, O. (1998) Self-interaction of

pneumolysin, the pore-forming protein toxin of *Streptococcus pneumoniae*. *Journal of Molecular Biology* **284**, 1223-1237.

Glatter, O. (1991) Small angle scattering and light scattering. In: Lindner, P. and Zemb, T.H., (Eds.) *Neutron, X-ray and Light Scattering: Introduction to an Investigative Tool for Colloidal and Polymeric Systems*, pp. 33-83. Elsevier Science (North Holland)

Gregory, L., Davis, K.G., Sheth, B., Boyd, J., Jefferis, R., Nave, C. and Burton, D.R. (1987) The solution conformations of the subclasses of human IgG deduced from sedimentation and small angle X-ray scattering studies. *Molecular Immunology* **24**, 821-829.

Grigsby, J.J., Blanch, H.W. and Prausnitz, J.M. (2000) Measurement of the rotational diffusion coefficient of lysozyme by depolarized light scattering: configuration of lysozyme in solution. *Journal of Physical Chemistry* **104**, 3645-3650.

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* **18**, 2714-2723.

Hantgan, R.R., Paumi, C., Rocco, M. and Weisel, J.W. (1999) Effects of ligand-mimetic peptides Arg-Gly-Asp-X (X = Phe, Trp, Ser) on α IIb β 3 integrin conformation and oligomerization. *Biochemistry* **38**, 14461-14474.

Harding, S.E. and Rowe, A.J. (1983) Modeling biological macromolecules in solution. II. The general triaxial ellipsoid. *Biopolymers* **22**, 1813-1829.

Harris, R.W., Sims, P.J. and Tweten, R.K. (1991) Evidence that *Clostridium perfringens* theta-toxin induces colloid-osmotic lysis of erythrocytes. *Infection and Immunity* **59**, 2499-2501.

Hildebrand, A., Pohl, M. and Bhakdi, S. (1991) *Staphylococcus aureus* α -toxin dual mechanism of binding to target cells. *Journal of Biological Chemistry* **266** 17195-17200.

Hoffmann, E.K. and Mills, J.W. (1999) Membrane events involved in volume regulation. *Current Topics in Membranes* **48**, 123-196.

- Huber, G.A. and McCammon, J.A. (1999) OOMPAA - Object-oriented model for probing assemblages of atoms. *Journal of Computational Physics* **151**, 264-282.
- Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33-38.
- Ilyina, E., Roongta, V., Pan, H., Woodward, C.K. and Mayo, K.H. (1997) A pulsed-field gradient NMR study of bovine pancreatic trypsin inhibitor self-association. *Biochemistry* **36**, 3383-3388.
- Iwamoto, M., Ohno-Iwashita, Y. and Ando, S. (1987) Role of the essential thiol group in the thiol-activated cytolysin from *Clostridium perfringens*. *European Journal of Biochemistry* **167**, 425-430.
- Jacrot, B. and Zaccai, G. (1981) Determination of molecular weight by neutron scattering. *Biopolymers* **20**, 2413-2426.
- Johnson, M.K. (1972) Properties of purified pneumococcal hemolysin. *Infection and Immunity* **6**, 755-760.
- Jones, D.T. (2000) Protein structure prediction in the postgenomic era. *Current Opinion in Structural Biology* **10**, 371-379.
- Jones, S. and Thornton, J.M. (1997) Prediction of protein-protein interaction sites using patch analysis. *Journal of Molecular Biology* **272**, 133-143.
- Kakalis, L.T. and Baianu, I.C. (1988) Oxygen-17 and deuterium nuclear magnetic relaxation studies of lysozyme hydration in solution: field dispersion, concentration, pH/pD, and protein activity dependences. *Archives of Biochemistry & Biophysics* **267**, 829-841.
- Kazenov, A.M., Maslova, M.N., Matskevich, Yu.A., Rustamov, A.M. and Shalabodov, A.D. (1998) Species variability of erythrocyte transport ATPases in mammals. *Comparative Biochemistry and Physiology* **119B**, 169-175.

Keefe, S.E. and Grant, E.H. (1974) Dipole moment and relaxation time of ribonuclease. *Physics in Medicine and Biology* **19**, 701-707.

Kim, H., Deonier, R.C. and Williams, J.W. (1977) The investigation of self - association reactions by equilibrium ultracentrifugation. *Chemical Reviews* **77**, 659-690.

Kirkwood, J.G. (1949) The statistical mechanical theory of irreversible processes in solutions of macromolecules (visco -elastic behaviour). *Journal of Chemistry and Physics* **68**, 649-660.

Kirkwood, J.G. (1954) The general theory of irreversible processes in solutions of macromolecules. *Journal of Polymer Science* **12**, 1-14.

Koenig, S.H. (1975) Brownian motion of an ellipsoid: A correction to Perrin's result. *Biopolymers* **14**, 2421-2423.

Korchev, Y.E., Bashford, C.L., Pederzolli, C., Pasternak, C.A., Morgan, P.J., Andrew, P.W. and Mitchell, T.J. (1998) A conserved tryptophan in pneumolysin is a determinant of the characteristics of channels formed by pneumolysin in cells and planar lipid bilayers. *Biochemical Journal* **329**, 571-577.

Krause, S. and O'Konski, C.T. (1963) Electric properties of macromolecules. VII. Kerr constants and rotational diffusion of some proteins in water and in glycerol-water solutions. *Biopolymers* **1**, 503-515.

Kumosinski, T.F. and Pessen, H. (1982) *Archives of Biochemistry & Biophysics* **219**, 89-100.

Kuntz, I.D. and Kauzmann, W. (1974) Hydration of proteins and polypeptides. In: Afsen, C.B., Edsall, J.T. and Richards, F.M., (Eds.) *Advances in Protein Chemistry*, pp. 239-345. Academic Press, San Fransisco.

Lakowicz, J.R., Laczko, G. and Gryczynski, I. (1987) Picosecond resolution of tyrosine fluorescence and anisotropy decays by 2 GHz frequency-domain fluorometry. *Biochemistry* **26**, 82-90.

Laskowski, R.A., MacArthur, M.W. and Thornton, J.M. (1998) Validation of protein models derived from experiment. *Current Opinion in Structural Biology* **8**, 631-639.

Laue, T.M. and Stafford, W.F. (1999) Modern applications of analytical ultracentrifugation. *Annual Review of Biophysics & Biomolecular Structure* **28**, 75-100.

Lawrence, A.J., Coote, J.G., MacLean, A.G., Parton, R. and Young, J.D. (1998) Conductimetry for enzyme teaching. *Biochemical Society Transactions* **26**, S197

Lawrence, A.J., Moores, G.R. and Steele, J. (1974) A conductimetric study of erythrocyte lysis by lysolecithin and linoleic acid. *European Journal of Biochemistry* **48**, 277-286.

Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology* **55**, 379-400.

Lundz-Katz, S., Laboda, H.M., McLean, L.R. and Phillips, M.C. (1988) Influence of molecular packing and phospholipid type on rates of cholesterol exchange. *Biochemistry* **27**, 3416-3423.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. and Huber, R. (1983) The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Journal of Molecular Biology* **39**, 480

Matadeen, R., Patwardhan, A., Gowen, B., Orlova, E., Pape, T., Cuff, M., Mueller, F., Brimacombe, R. and van Heel, M. (1999) The *Escherichia coli* large ribosomal subunit at 7.5 Å resolution. *Structure* **7**, 1575-1583.

Mayans, M.O., Coadwell, W.J., Beale, D., Symons, D.B.A. and Perkins, S.J. (1995) Demonstration by pulsed neutron scattering that the arrangement of the Fab and Fc fragments in the overall structures of bovine IgG1 and IgG2 in solution is similar. *Biochemical Journal* **311**, 283-281.

McRorie, D.K. and Voelker, P.J. (1993) *Self-Associating Systems in the Analytical Ultracentrifuge*, California: Beckman Instruments.

Media Selection Guide (1996) POROS: Columns and Media Selection Guide. 8-1119-10-0294. USA: Perseptive Biosystems.

Mezna, M. and Lawrence, A.J. (1994) Conductimetric assays for the hydrolase and transferase activities of phospholipase D enzymes. *Analytical Biochemistry* **218**, 370-376.

Milner-White, E.J., Belhadjmostefa, K. and Poet, R. (1990) Computer-generated diagrams for relating primary structure to 3-dimensional features in proteins. *Biochemical Society Transactions* **18**, 917-918.

Minton, A.P. (1989) Lateral diffusion of membrane proteins in protein-rich membranes. A simple hard particle model for concentration dependence of the two dimensional diffusion coefficient. *Biophysical Journal* **55**, 805-808.

Mitchell, T.J., Walker, J.A., Saunders, F.K., Andrew, P.W. and Boulnois, G.J. (1989) Expression of the pneumolysin gene in *Escherichia coli*: rapid purification and biological properties. *Biochimica et Biophysica Acta* 67-72.

Mooney, R.A. (1988) Use of digitonin-permeabilised adipocytes for cAMP studies. *Methods in Enzymology* **18**, 193-202.

Morgan, P.J., Hyman, S.C., Byron, O., Andrew, P.W., Mitchell, T.J. and Rowe, A.J. (1994) Modelling the bacterial protein toxin, pneumolysin, in its monomeric and oligomeric form. *Journal of Biological Chemistry* **269**, 25315-25320.

Morgan, P.J., Hyman, S.C., Rowe, A.J., Mitchell, T.J., Andrew, P.W. and Saibil, H.R. (1995) Subunit organisation and symmetry of pore - forming oligomeric pneumolysin. *FEBS Letters* **371**, 77-80.

Morgan, P.J., Varley, P.G., Rowe, A.J., Andrew, P.W. and Mitchell, T.J. (1993) Characterization of the solution properties and conformation of pneumolysin, the membrane - damaging toxin of *Streptococcus pneumoniae*. *Biochemical Journal* **296**, 671-674.

Mosimann, S., Meleshko, R. and James, M.N.G. (1995) A critical assessment of comparative molecular modelling of tertiary structures of proteins. *Proteins* **23**, 301-317.

Mueller, F., Sommer, I., Baranov, P., Matadeen, R., Stoldt, M., Wohnert, J., Gorlach, M., van Heel, M. and Brimacombe, R. (2000) The 3D arrangement of the 23 S and 5 S rRNA in the *Escherichia coli* 50 S ribosomal subunit based on a cryo-electron microscopy reconstruction at 7.5 Å resolution. *Journal of Molecular Biology* **298**, 35-59.

Nelson, G.J. (1967) Lipid composition of erythrocytes in various mammalian species. *Biochimica et Biophysica Acta* **144**, 221-232.

Nicoli, D.F. and Benedek, G.B. (1976) Study of thermal denaturation of lysozyme and other globular proteins by light scattering spectroscopy. *Biopolymers* **15**, 2421-2437.

Osborne, J.C., Miller, J.H. and Kempner, E.S. (2000) Molecular mass and volume in radiation target theory. *Biophysical Journal* **78**, 1698-1702.

Ostolaza, H., Bakas, L. and Goni, F.M. (1997) Balance of electrostatic and hydrophobic interactions in the lysis of model membranes by *E. coli* α haemolysin. *Journal of Membrane Biology* **158**, 137-145.

Otting, G., Liepinsh, E. and Wuthrich, K. (1991) Protein hydration in aqueous solution. *Science* **254**, 974-980.

Paci, E. and Velikson, B. (1997) On the volume of macromolecules. *Biopolymers* **41**, 785-797.

Perkins, S.J. (1986) Protein volumes and hydration effects: The calculation of partial specific volumes, neutron scattering matchpoints and 280 nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *European Journal of Biochemistry* **157**, 169-180.

Perkins, S.J. (1988a) Structural studies of proteins by high-flux X-ray and neutron solution scattering. *Biochemical Journal* **245**, 313-327.

Perkins, S.J. (1988b) X-ray and neutron solution scattering. In: Neuberger, A. and Deenen, L.L.M.V., (Eds.) *Modern Physical Methods in Biochemistry*, Elsevier Science Publishers, Amsterdam, New York, Oxford.

Perkins, S.J. (1989) Hydrodynamic modelling of complement. In: Harding, S.E. and Rowe, A.J., (Eds.) *Dynamic properties of biomolecular assemblies*, pp. 226-245. Royal Society of Chemistry, Cambridge, UK.

Perkins, S.J., Ashton, A.W., Boehm, M.K. and Chamberlain, D. (1998) Molecular structures from low angle X-ray and neutron scattering studies. *International Journal of Biological Macromolecules* **22**, 1-16.

Perrin, F. (1934) Mouvement brownien d'un ellipsoïde. II. Rotation libre et depolarisation des fluorescences. Translation et diffusion de molécules ellipsoïdales. *J. Phys. Radium*. **5**, 497-511.

Ralston, G. (1993) *Introduction to Analytical Ultracentrifugation*, Beckman Instruments, Inc., Palo Alto, CA, USA.

Rimai, L., Hickmott, J.T., Cole, T. and Carew, E.W. (1970) Quasi-elastic light scattering by diffusional fluctuations in RNase solutions. *Biophysical Journal* **10**, 20-37.

Rossjohn, J., Gilbert, R.C., Crane, D., Morgan, P.J., Mitchell, T.J., Rowe, A.J., Andrew, P.W., Paton, J.C., Tweten, R.K. and Parker, M.W. (1998) The molecular mechanism of pneumolysin, a virulence factor from *Streptococcus pneumoniae*. *Journal of Molecular Biology* **284**, 449-461.

Rostovtseva, T.K., Bashford, C.L., Alder, G.M., Hill, G.N., McGiffert, C., Apel, P.Y., Lowe, G. and Pasternak, C.A. (1996) Diffusion through narrow pores - movement of ions, water and nonelectrolytes through track-etched PETP membranes. *Journal of Membrane Biology* **151**, 29-43.

Rothen, A. (1940) Molecular weight and electrophoresis of crystalline ribonuclease. *Journal of General Physiology* **24**, 203-211.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, second edn. Cold Spring Harbor Laboratory Press.

Sayle, R.A. and Milner-White, E.J. (1995) RASMOL -Biomolecular graphics for all. *Trends in Biochemical Sciences* **20**, 374-376.

Scheraga, H.A. and Mandelkern, L. (1953) Consideration of the hydrodynamic properties of proteins. *Journal of the American Chemical Society* **75**, 179-184.

Seigneuret, M. and Devaux, P.F. (1984) ATP-dependent asymmetric distribution of spin-labelled phospholipids. *Proceedings of the National Academy of Sciences (USA)* **81**, 3751-3755.

Sheehan, D. (2000) Chromatography. In: Anonymous *Physical Biochemistry*, 1 edn. 12-59.

Simha, R. (1940) *Journal of Physical Chemistry* **44**, 25

Sjoberg, B. and Mortensen, K. (1997) Structure and thermodynamics of nonideal solutions of colloidal particles: Investigation of salt-free solutions of human serum albumin by using small-angle neutron scattering and Monte Carlo simulation. *Biophysical Chemistry* **65**, 75-83.

Sophianopolous, A.J., Rhodes, C.K., Holcomb, D.N. and van Holde, K.E. (1962) Physical studies of lysozyme. I. Characterisation. *Journal of Biological Chemistry* **237**, 1107-1112.

Spotorno, B., Piccinini, L., Tassara, G., Ruggiero, C., Nardini, M., Molina, F. and Rocco, M. (1997) BEAMS (BEAd Modelling System): A set of computer programs for the generation, the visualization and the computation of the hydrodynamic and conformational properties of bead models of proteins. *European Biophysics Journal* **25**, 373-384.

Steiner, R.F. and McAllister, A.J. (1957) Use of the fluorescence technique as an absolute method for obtaining mean relaxation times of globular proteins. *Journal of Polymer Science* **24**, 105-123.

Stroustrup, B. (1995) *The C++ Programming Language*, 2 edn. Addison-Wesley Publishing Company.

Svergun, D.I. (1992) Determination of the regularisation parameter in indirect-transform methods using perceptual criteria. *Journal Of Applied Crystallography* **25**, 495-503.

Svergun, D.I., Barberato, C. and Koch, M.H.J. (1995) CRY SOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773.

Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal* **76**, 2879-2886.

Svergun, D.I., Barberato, C., Koch, M.H.J., Fetler, L. and Vachette, P. (1997) Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins: Structure, Function, & Genetics* **27**, 110-117.

Svergun, D.I., Petoukhov, M.V., Koch, M.H.J. and Konig, S. (2000) Crystal versus solution structures of thiamine diphosphate-dependent enzymes. *Journal of Biological Chemistry* **275**, 297-302.

Swanson, E., Teller, D.C. and de Haën, C. (1978) The low Reynolds number translational friction of ellipsoids, cylinders, dumbbells, and hollow spherical caps. Numerical testing of the validity of the modified Oseen tensor in computing the friction of objects modeled as beads on a shell. *Journal of Chemical Physics* **68**, 5097-5102.

Tanford, C. (1961) *Physical Chemistry of Macromolecules*, New York: John Wiley and Sons.

Tao, Y. and Zhang, W. (2000) Recent developments in cryo-electron microscopy reconstruction of single particles. *Current Opinion in Structural Biology* **10**, 616-622.

Teller, D.C., Swanson, E. and de Haën, C. (1979) The translational frictional coefficients of proteins. *Methods in Enzymology* **61** 103-124.

Thanki, N., Thornton, J.M. and Goodfellow, J.M. (1988) Distributions of water around amino acid residues in proteins. *Journal of Molecular Biology* **202**, 637-657.

Tilton, R.F., Dewan, J.C. and Petsko, G.A. (1992) Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease A at nine different temperatures from 98 to 320 K. *Biochemistry* **31**, 2469

Tsai, J., Taylor, R., Chothia, C. and Gerstein, M. (1999) The packing density in proteins: Standard radii and volumes. *Journal of Molecular Biology* **290**, 253-266.

van Holde, K.E. (1985) Solutions of Macromolecules. In: Anonymous *Physical Biochemistry*, 2 edn. pp. 24-50. Prentice-Hall, Englewood Cliffs, New Jersey.

Walker, J.A., Allen, R.L., Falmagne, P., Johnson, M.K. and Boulnois, G.J. (1987) Molecular cloning, characterization, an complete nucleotide sequence of the gene for pneumolysin, the sulphhydryl-activated toxin of *Streptococcus pneumoniae*. *Infection & Immunity* **55**, 1184-1189.

Wang, P.-F., Veine, D.M., Ahn, S.H. and Williams, C.H. (1996) A stable mixed disulfide between thioredoxin reductase and its substrate, thioredoxin: preparation and charachterization. *Biochemistry* **35**, 4812-4819.

Webster, D.M. (2000) *Protein Structure Prediction: Methods and Protocols*, Humana Press.

Weik, M., Ravelli, R.B.G., Ryger, G., Sweeney, S., Aves, M.L., Arel, M., Ros, P., Ilman, I., Roon, J. and Ussman, J.L. (2000) Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proceedings of the National Academy of Sciences (USA)* **97**, 623-628.

Williams, M.A., Goodfellow, J.M. and Thornton, J.M. (1994) Buried waters and internal cavities in monomeric proteins. *Protein Science* **3**, 1224-1235.

Williams, R.J.P. (1974) Calcium ions: their ligands and their functions. *Biochemistry Society Symposia* **39**, 133-138.

Wills, P.R. and Georgalis, Y. (1981) Concentration dependence of the diffusion coefficient of a dimerising protein: bovine pancreatic trypsin inhibitor. *Journal of Physical Chemistry* **85**, 3976-3984.

Wilson, K.P., Malcolm, B.A. and Matthews, B.W. (1993) Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *Journal of Biological Chemistry* **267**, 10842-10849.

Wriggers, W., Milligan, R.A., Schulten, K. and McCammon, J.A. (1998) Self-organizing neural networks bridge the biomolecular resolution gap. *Journal of Molecular Biology* **284**, 1247-1254.

Xenarios, I., Rice, D.W., Salwinski, L., Baron, M.K., Marcotte, E.M. and Eisenberg, D. (2000) DIP: the database of interacting proteins. *Nucleic Acids Research* **28**, 289-291.

Zamyatnin, A.A. (1972) Protein volume in solution. *Progress In Biophysics & Molecular Biology* **24**, 109-123.

Zipper, P. and Durchschlag, H. (1998) Recent advances in the calculation of hydrodynamic parameters from crystallographic data by multibody approaches. *Biochemical Society Transactions* **26**, 726-731.

Zipper, P. and Durchschlag, H. (2000) Prediction of hydrodynamic and small angle scattering parameters from crystal and electron microscopic structures. *Journal of Applied Crystallography* **33**, 788-792.

Zwaal, R.F.A., Roelofsen, B., Comfurius, P. and van Deenen, L.L.M. (1975) Organization of phospholipids in human red cell membranes as detected by the action of various purified phospholipases. *Biochimica et Biophysica Acta* **406**, 83-96.

Zwanzig, R., Kiefer, J. and Weiss, G.H. (1968) On the validity of the Kirkwood-Riseman Theory. *Proceedings of the National Academy of Sciences (USA)* **60**, 381-386.

APPENDIX 1

Name	Description
Agg	Aggregate class contains objects for storing and relating different types of information (objects).
Vector	Used for when vector calculations are required
Timer	Sets timing events
Graph	Calls graphics routines
Grid	Generates a virtual cubic grid.
Protein	Stores information about a molecule in terms of atoms, residues and subunits.
Win	Creates windows
Gui	
Draw	Draws graphics windows and buttons.

